Bayesian Analysis for Time Series of Count Data

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Doctor of Philosophy

in the Department of Mathematics and Statistics

University of Saskatchewan

Saskatoon

By

Mohammed Obeidat

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
142 McLean Hall
106 Wiggins Road
University of Saskatchewan
Saskatoon, SK S7N 5E6
CANADA

Abstract

Time series involving count data are present in a wide variety of applications. In many applications, the observed counts are usually small and dependent. Failure to take these facts into account can lead to misleading inferences and may detect false relationships. To tackle such issues, a Poisson parameter-driven model is assumed for the time series at hand. This model can account for the time dependence between observations through introducing an autoregressive latent process.

In this thesis, we consider Bayesian approaches for estimating the Poisson parameter-driven model. The main challenge is that the likelihood function for the observed counts involves a high dimensional integral after integrating out the latent variables. The main contributions of this thesis are threefold. First, I develop a new single-move (SM) Markov chain Monte Carlo (MCMC) method to sample the latent variables one by one. Second, I adopt the idea of the particle Gibbs sampler (PGS) method (Andrieu et al., 2010) into our model setting and compare its performance with the SM method. Third, I consider Bayesian composite likelihood methods and compare three different adjustment methods with the unadjusted method and the SM method. The comparisons provide a practical guide to what method to use.

We conduct simulation studies to compare the latter two methods with the SM method. We conclude that the SM method outperforms the PGS method for small sample size, while they perform almost the same for large sample size. However, the SM method is much faster than the PGS method. The adjusted Bayesian composite methods provide closer results to the SM than the unadjusted one. The PGS and the selected adjustment method from simulation studies are compared with the SM method via a real data example. Similar results are obtained: first, the PGS method provides results very close to those of the SM method. Second, the adjusted composite likelihood methods provide closer results to the SM than the unadjusted one.

Acknowledgements

I would like to express my sincere thanks to my supervisor, Dr. Juxin Liu. I really learned a lot under her supervision. Throughout my research, I found her very helpful and was always available whenever I needed help. This work could not be possible without her support and encouragement.

I am also very grateful to all the committee members: Prof. Chris Soteros, Prof. Mik Bickis, Dr. Nathaniel Osgood, and Dr. Longhai Li for sharing their invaluable comments and suggestions, which made the presentation of the thesis more clear. Special thanks to Dr. Longhai Li for the permission to use his cluster to do the simulation studies. I would like also to thank Dr. Shahedul Khan for his help during the early stages of writing my C codes. Thanks to Prof. Harry Joe for the permission to use the real dataset.

Most importantly, I am deeply indebted to my family, who missed me back home and counted down every single day throughout the period of study. Special thanks to my wife, Rasha, for staying with me and for her continuing love, support, and encouragement throughout this period. I also thank my sons, Ahmad, Jad and Abdelrahman and my daughter, Kinda, for all the happy moments I spent with them in my spare time. They always made me happy and put a smile in my face.

The work in this thesis was funded by the College of Graduate Studies and Research, Department of Mathematics and Statistics and Dr. Juxin Liu, University of Saskatchewan, whom I gratefully acknowledge.

Mohammed Obeidat

Saskatoon, 2014

To my parents, my wife, and those who educated me.

Contents

Permission to Use i

Abstract ii

Acknowledgements iii

Contents v

List of Tables vii

List of Figures ix

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Parameter-driven Model for Poisson Counts . . . 2
- 1.3 Literature Review on Methods for Analyzing Time Series of Counts . . . 3
- 1.4 Contributions and Outline of the Thesis . . . 5

2 Single Move Algorithm 7
- 2.1 Bayesian Approach . . . 7
  - 2.1.1 The Metropolis-Hastings Algorithm . . . 8
- 2.2 The Posterior Distribution of the Model Parameters . . . 9
  - 2.2.1 Prior Distributions . . . 11
- 2.3 The Single Move Method . . . 12
  - 2.3.1 Sampling the Latent Process . . . 14
  - 2.3.2 Sampling the Poisson Regression Parameters $\boldsymbol{\beta}$ . . . 16
  - 2.3.3 Sampling $\{r_1^*, \ldots, r_p^*\}$ . . . 17
  - 2.3.4 Sampling σ_V^2 . . . 18

3 The Particle Gibbs Sampler 19
- 3.1 Motivation . . . 19
- 3.2 Sequential Monte Carlo Methods . . . 20
- 3.3 The Particle Gibbs Sampler . . . 23
  - 3.3.1 Illustrative Example . . . 24
  - 3.3.2 Resampling Step . . . 25
  - 3.3.3 The Proposal Density . . . 27
  - 3.3.4 The Number of Particles . . . 28
- 3.4 Simulation Study Design and Results . . . 30

4 Bayesian Estimation Using Composite Likelihood 40
- 4.1 Motivation . . . 40

4.2 Composite Likelihood . . . 42
4.3 Adjustment of The Composite Likelihood . . . 45
4.3.1 Magnitude Adjustment . . . 45
4.3.2 Curvature Adjustment . . . 46
4.3.3 Open Faced Sandwich Adjustment . . . 47
4.3.4 Estimating H and J . . . 48
4.4 MCMC Using Composite Likelihood . . . 49
4.4.1 Adjusted Metropolis Hastings Algorithm . . . 49
4.4.2 Adjusted Gibbs Sampler . . . 50
4.5 Poisson Parameter Driven Model . . . 52
4.5.1 The Metropolis-Hastings Algorithm . . . 52
4.5.2 Gradients of the Log Composite Likelihood . . . 53
4.5.3 Evaluating the Composite Likelihood and the Gradients . . . 54
4.6 Simulation Study Results . . . 56

5 Analysis of Time Series of Car Crash Data **65**
5.1 Data Description . . . 65
5.2 Analysis and Results . . . 66
5.3 Model Adequacy Check . . . 68

6 Discussion and Future Plan **74**
6.1 Main Results . . . 74
6.2 Future Work . . . 75

References **77**

A Complete Results for Chapter 3 **82**

B Complete Results for Chapter 4 **89**

C C Codes for The SM Method **103**

List of Tables

5.1 Posterior Summary of Parameters Using Methods discussed in the Thesis . . 68

B.2 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 90

B.3 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 91

B.4 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 91

B.5 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 92

B.6 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 92

B.7 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 93

B.8 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 93

B.9 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$. 94

B.10 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.20, 0.10, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$. 94

B.11 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.40, 0.20, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 95
B.12 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.88, 0.34, −0.28) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 96
B.13 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.20, 0.10, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 97
B.14 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.40, 0.20, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 98
B.15 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.88, 0.34, −0.28) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 99
B.16 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.20, 0.10, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 100
B.17 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.40, 0.20, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 101
B.18 Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.88, 0.34, −0.28) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$. 102

List of Figures

3.1 The structure of the conditional SMC 25
3.2 The selected path 26
3.3 The posterior and proposal densities 29
3.4 Percentage of times of selecting an independent particle 31
3.5 ACF plot for posterior samples using SM and PGS methods 34
3.6 RMSE and relative bias of parameter estimates for the SM and the PGS methods with AR(2) latent process and weak serial correlation 37
3.7 RMSE and relative bias of parameter estimates for the SM and the PGS methods with AR(2) latent process and moderate serial correlation 38
3.8 RMSE and relative bias of parameter estimates for the SM and the PGS methods with AR(2) latent process and strong serial correlation 39

4.1 RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and weak correlation structure 59
4.2 Bias of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and weak correlation structure 60
4.3 RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and moderate correlation structure 61
4.4 Bias of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and moderate correlation structure 62
4.5 RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and strong correlation structure 63
4.6 Bias of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process and strong correlation structure 64

5.1 Normal qq-plot of ordinary pseudo-residuals and forecast pseudo-residuals for car crashes data. 73

Chapter 1

Introduction

1.1 Background

Time series involving count data are present in a wide variety of applications. Examples can be easily found in epidemiology, finance, accident prevention and many other research areas. In epidemiology, for example, time series of count data can be seen in disease surveillance where the occurrence of rare infectious disease is often monitored by health officials to study the changes in patterns of disease. The occurrence of the disease is often expressed as daily, weekly or monthly counts.

The majority of the applications mentioned above assume Poisson distribution for the time series of counts. This is because the events of interest are relatively rare, and the time scale on which the measurements observed is small, which may lead to small counts to be observed. When the counts are relatively large, a Gaussian assumption of the counts is convenient (Cameron and Trivedi, 1998). However, relatively rare events make the assumption of the Gaussian distribution questionable. Expanding the time scale in which the data are collected to get larger counts to justify the use of Gaussian distribution would eliminate or, at least, distort the relationship between the explanatory variables and the response variable. For example, in a study of the relationship between weather conditions and number of car accidents, Brijs et al. (2008) worked with daily measurements of weather conditions and car accidents counts. Expanding the time scale in this example to weekly or monthly instead of daily could be misleading, as the weather covariates may change noticeably during the days of the week.

Another fact regarding time series is that temporal correlation is often present between adjacent observations. To account for this temporal dependence in the data, two types of Poisson models have been proposed in the literature (Cox, 1981): parameter-driven and observation-driven models. In observation-driven models, the conditional mean of the current observation, given past observations, is determined as a function of these past observations. In parameter-driven models, the dependence between observations is addressed by introducing a latent process. In this model, it is assumed that the observations are correlated marginally but independent conditional on this latent process. The advantage of the Poisson parameter-driven model over the observation-driven model is the ease of interpretation of the regression parameters.

Failure to take the non-Gaussian distribution of the counts and temporal dependence into account may lead to false inference and can detect wrong relationships. In this thesis, I tackle the above mentioned issues by assuming a Poisson parameter-driven model for the time series at hand. The inference of the Poisson parameter-driven model is challenging because the likelihood function involves a high dimensional integral over the joint distribution of the latent process.

In this thesis, I discuss Bayesian estimation of the Poisson parameter-driven model with autoregressive $AR(p)$ latent process. Three Bayesian approaches are developed: the single-move (SM) method, the particle Gibbs sampler (PGS) method, and composite likelihood methods with and without adjustments.

1.2 Parameter-driven Model for Poisson Counts

Let $Y_1, \ldots, Y_n$ be a sequence of counts for occurrence of an event of interest that are observed at equally spaced time points. In the Poisson parameter-driven model, it is assumed that $Y_t|\lambda_t \sim \text{Poisson}(\lambda_t)$, where λ_t depends on unobserved time-varying variable η_t and a $(K+1)$-dimensional vector of covariates X_t, through model parameter vector $\boldsymbol{\beta}$. The first element of X_t is 1 for the intercept and the remaining elements correspond to the values of the K

covariates at time t. Then

$$Y_t|\lambda_t \sim \text{Poisson}\left(\lambda_t\right), \quad t = 1, \ldots, n, \tag{1.1}$$

$$\lambda_t = \exp\left\{\boldsymbol{\beta}^T X_t + \eta_t\right\}, \tag{1.2}$$

The time dependence in Y_t is embedded in the latent variables $\{\eta_t\}$, which are modeled by a Gaussian autoregressive process of order p, denoted by AR(p), for a small positive integer p, as follows:

$$\eta_t = \phi_1\eta_{t-1} + \phi_2\eta_{t-2} + \cdots + \phi_p\eta_{t-p} + V_t, \quad t > p, \tag{1.3}$$

where $\{V_t\}$, $t = 1, \ldots, n$ are iid normal random variables with mean zero and variance σ_V^2.

The likelihood function for the observed data Y_t involves high dimensional integral with the latent process integrated out

$$h\left(\mathbf{Y}\right) = \int \ldots \int \left\{\prod_{t=1}^{n} g\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)\right\} f\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_V^2\right) d\eta_1 \ldots d\eta_n, \tag{1.4}$$

where g is the Poisson probability mass function, and f is the multivariate Gaussian density function. The high dimensional integral in the likelihood function cannot be solved in closed form. In this thesis, I consider Bayesian approaches to deal with this issue.

1.3 Literature Review on Methods for Analyzing Time Series of Counts

This section reviews the available literature on analyzing Poisson parameter-driven models for a single time series of count data in both frequentist and Bayesian paradigm. The first researcher to discuss the Poisson parameter-driven model in the frequentist paradigm was Zeger (1988). Zeger employed a generalized estimating equations approach to estimate the unknown parameters of the Poisson parameter-driven model. An advantage of Zeger's approach is that it does not specify any distributional assumption on the latent process and only first-order and second-order moments need to be specified. On the contrary, for higher-order autoregressive models, this approach may not yield admissible parameter estimates or even

lead to situations where the equations do not have solutions (Ng et al., 2011). To improve efficiency of the estimators, a parametric distribution on the latent process needs to be assumed. This assumption results in a complex likelihood which makes the estimation procedures more complicated than what is discussed by Zeger. The main challenge is how to evaluate the high dimensional integral involved in the likelihood function as it cannot often be evaluated analytically, and the computation is prohibitive. There has been plentiful literature about how to tackle this challenging issue. One popular approach is called the simulated maximum likelihood method (Jung et al., 2006; Richard and Zhang, 2007) and (Chan and Ledolter, 1995; McCulloch, 1997; Kuk and Cheng, 1997), which estimates the above integral by the Monte Carlo (MC) technique. The first group of authors used importance sampling methods, while the second used the Monte Carlo Expectation Maximization algorithm (MCEM). The main emphasis of the work is how to develop the MC sampler (or importance sampler in importance sampling) so as to provide an accurate and efficient estimate of the parameters of interest. Another simulation-based method is approximate Bayesian computation (Pritchard et al., 1999; Marjoram et al., 2003) that bypasses the evaluation of the likelihood function. The basic idea is to 'accept' a parameter value if the simulated data set conditional on this parameter value is *similar* to the actual observed data. This method is most suitable for cases where the simulation is straightforward while the likelihood is difficult –or even impossible– to compute. Different from the above numerical methods, the composite likelihood method (Lindsay, 1988) replaces the high-dimensional integral by lower-dimensional integrals. This method has rapidly gained popularity in the past decades. For a recent review on this topic, please see Varin et al. (2011). Recently, Ng et al. (2011) and Davis and Yau (2011) discussed the use of composite likelihood in Poisson parameter-driven models within the frequentist paradigm. More information on the composite likelihood methods can be found in Section 4.1 and Section 4.2.

Our objective here is to develop Bayesian approaches for Poisson parameter-driven models based on the *full* and *composite* likelihood functions, respectively. A substantial body of literature investigates time series data models analyzed in the Bayesian paradigm. Durbin

and Koopman (2000) used importance sampling approach to perform full Bayesian analysis. The major challenge is the sampling of the latent variables. Oh and Lim (2001) used a Bayesian MCMC method for estimating a Poisson parameter-driven model with an $AR(1)$ latent process. They reformulate the Poisson parameter-driven model in terms of extra normally distributed latent variables and then approximate the Poisson cumulative distribution function (cdf) by a normal cdf. By so doing, they were able to sample the latent variables one by one from a normal distribution that is restricted to a fixed region. However, sampling these variables one at a time using MCMC methods tends to be inefficient (Shephard and Pitt, 1997). Jung et al. (2006) discuss how to sample the latent variables from their joint posterior distribution using a combination of the efficient importance sampler (EIS) with accept-reject Metropolis Hastings (AR-MH). The first of these was developed by Richard and Zhang (2007) and the latter by Tierney (1994). Fruhwirth-Schnatter and Wagner (2006) and Fruhwirth-Schnatter et al. (2009) used data augmentation and MCMC methods. They developed an auxiliary mixture sampler to sample all latent variables at once from a multivariate normal density. They use a multivariate normal distribution to approximate the posterior density by introducing the arrival and inter-arrival time variables associated with the Poisson counts and then approximate their distributions by a mixture of normal distributions.

1.4 Contributions and Outline of the Thesis

In the existing literature, most of the work in Bayesian analysis of Poisson parameter-driven model either assumes $AR(1)$ latent process or assumes a general $AR(p)$ process but does not directly sample from the posterior distribution. For example, Oh and Lim (2001) and Jung et al. (2006) assume $AR(1)$ latent process, and Durbin and Koopman (2000), Fruhwirth-Schnatter and Wagner (2006) and Fruhwirth-Schnatter et al. (2009) assume $AR(p)$ latent process, but did not directly sample from the posterior of the latent process. In this thesis, I develop Bayesian approaches to analyzing a single time series of counts in the context of Poisson parameter-driven model with an $AR(p)$ latent process. This model belongs to the

class of non-linear state-space models (SSM) with non-Gaussian response variables.

The main contributions of the thesis are three-fold: first, in Chapter 2, I develop a new single-move (SM) method to sample the latent variables one at a time using the accept-reject method. Second, with a hope of improving the efficiency of the MCMC algorithm, I propose to use the particle Gibbs sampler (PGS) method to sample the latent variables as a block and compare its performance with the SM method. The PGS was developed by Andrieu et al. (2010) to sample from any high-dimensional density, however, to our best knowledge, it has not yet been discussed in the context of Poisson parameter-driven models. This was discussed in Chapter 3. The chapter concludes with simulation studies to compare the performance of the PGS method with the SM method. Third, in Chapter 4, I consider Bayesian composite likelihood methods to analyze Poisson parameter-driven models, and compare their performance with the SM method. For the Bayesian composite likelihood methods to provide appropriate inference, adjustment of the composite likelihood when being used in the posterior formula, in place of the full likelihood, needs to be taken into account. Three different adjustment methods that have recently been proposed in the literature are discussed and compared. Chapter 4 concludes with simulation studies to compare the performance of the adjusted and unadjusted composite likelihood methods with the SM method. All of these methods are compared with the SM method in terms of bias and root mean square error (RMSE) through simulation studies. More details about the comparisons results can be found in Section 3.4, Section 4.6 and Section 6.1.

In Chapter 5, I analyze a real data example that consists of daily car crashes on a major road near Schiphol, the largest airport in the Netherlands, in 2001. The SM, PGS methods and the adjustment method that perform the best in simulation studies are used to analyze the real data example. Chapter 6 concludes with findings and future plans.

Chapter 2

Single Move Algorithm

In this chapter, we propose a new single-move MCMC algorithm. The 'single-move' designation refers to the one-by-one update of the latent variables. The full conditionals were written in a convenient way for the SM method. For parameters whose full conditionals from which it is not easy to directly sample, we used either the accept-reject method or the Metropolis-Hastings (M-H) method.

2.1 Bayesian Approach

In the Bayesian paradigm, the model parameters, $\boldsymbol{\theta}$, are treated as being random variables and a distribution is assigned to these parameters. This distribution is called the prior distribution and is denoted by $\pi(\boldsymbol{\theta})$. The prior distribution reflects the researcher's knowledge of the parameters prior to collecting the data. Bayesian inference relies on the posterior distribution, the distribution of the parameters given the data $\mathbf{y}$, denoted by $\pi(\boldsymbol{\theta}|\mathbf{y})$. The posterior distribution of the parameters given the data can be obtained by applying Bayes' theorem, which is accomplished simply by multiplying the prior by the probability density of the data given the parameter, $f(\mathbf{y}|\boldsymbol{\theta})$, then normalizing.

$$\pi(\boldsymbol{\theta}|y) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\int \ldots \int f(\mathbf{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}. \tag{2.1}$$

The normalizing constant in Eq. (2.1) above cannot usually be evaluated; the posterior is thus commonly written as follows:

$$\pi(\boldsymbol{\theta}|y) \propto f(\mathbf{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}).$$

Bayesian inference is based on the posterior distribution. For example, if one wishes to find the Bayes' estimator of a parameter θ_i, the i^{th} element of $\boldsymbol{\theta}$, then he/she has to specify a loss function, and determine the value of θ_i that minimizes the posterior expected loss. This value will be the Bayes' estimator of the parameter θ_i. One popular choice of the loss function is the squared-error loss function. The Bayes estimator of the parameter θ_i under the squared-error loss function is the posterior mean; that is

$$\hat{\theta}_i = \int \cdots \int \theta_i \pi\left(\boldsymbol{\theta}|y\right) d\boldsymbol{\theta}.$$

In many modern applications, it is hard to evaluate the integral in the above equation analytically. Therefore, it may be necessary to turn to simulation based methods to approximate the above integral. MCMC methods represent a popular tool for this purpose. Two known MCMC methods are the Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970) and the Gibbs sampler (Geman and Geman, 1984). These methods simulate slightly dependent draws. After running the algorithm long enough, these draws can be considered to be from the posterior distribution of interest. These draws can then be used to approximate posterior quantities or probabilities. For example, the posterior mean can be approximated by the empirical mean of the draws.

2.1.1 The Metropolis-Hastings Algorithm

The M-H algorithm is one of the MCMC methods that can be used to draw samples from any posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$ that is known up to a normalizing constant, that is

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\gamma(\boldsymbol{\theta}, \mathbf{y})}{K},$$

where $\gamma(\boldsymbol{\theta}, \mathbf{y}) = f\left(\mathbf{y}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)$, and K is the normalizing constant, which usually cannot be evaluated in practice.

The M-H algorithm generates samples iteratively in such a way that the distribution of the next sample depends only on the current one. The distribution of these samples approximates the target density increasingly closely as the number of samples rises. To get M draws from $\pi(\boldsymbol{\theta}|\mathbf{y})$, the M-H algorithm follows the following steps:

1. Choose a starting value $\boldsymbol{\theta}^{(0)}$.

2. Repeat the following steps for $i = 1, \ldots, M$:

   - At iteration i, draw a candidate $\boldsymbol{\theta}^*$ from a proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$.
   - Compute an acceptance ratio

$$r = \frac{\gamma(\boldsymbol{\theta}^*, \mathbf{y}) q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}{\gamma(\boldsymbol{\theta}^{(i-1)}, \mathbf{y}) q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})}.$$

   - Sample u from the uniform distribution $\mathrm{U}(0, 1)$.
   - Accept $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^{(i)}$ if $u < min(r, 1)$. Otherwise, let $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$.

The samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}$ can be used to obtain posterior summaries.

2.2 The Posterior Distribution of the Model Parameters

In the Poisson parameter driven model (1.1)-(1.3), the parameters are $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and σ_V^2. From now on, the vector $\boldsymbol{\theta}$ will be used to represent $\{\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma_V^2\}$ when needed. The posterior distribution of the parameters and the latent process is given by

$$\pi\left(\boldsymbol{\theta}, \boldsymbol{\eta}|y\right) \propto \left\{ \prod_{t=1}^{n} g\left(y_t \mid \eta_t, \boldsymbol{\beta}\right) \right\} f_{\boldsymbol{\eta}}\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_V^2\right) \pi_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}\right) \pi_{\boldsymbol{\phi}}\left(\boldsymbol{\phi}\right) \pi_{\sigma_V^2}\left(\sigma_V^2\right), \tag{2.2}$$

where g is the Poisson probability mass function, $f_{\boldsymbol{\eta}}$ is the joint density of the latent variables, and $\pi_{\boldsymbol{\beta}}$, $\pi_{\boldsymbol{\phi}}$, $\pi_{\sigma_V^2}$ are the prior densities on $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and σ_V^2, respectively. More details on how we selected the prior densities of the parameters are given in Section 2.2.1.

In Eq. (2.2) above, the joint distribution of the latent process, $f_{\boldsymbol{\eta}}\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_V^2\right)$, can be obtained as follows. First, notice that the first p variables of the process, $\eta_1, \ldots, \eta_p$, in Eq. (1.3) are drawn from a multivariate Gaussian distribution with mean 0 and covariance matrix Σ, and V_t are independent Gaussian random variables with mean zero and variance σ_V^2. The AR(p) process in Eq. (1.3) is said to be stationary if the following are satisfied (Brockwell and Davis (1991), page 12).

- $\mathrm{E}\left(|\eta_t|^2\right) < \infty$ for all t;
- $\mathrm{E}\left(\eta_t\right) = m$ for all t;
- $\mathrm{Cov}\left(\eta_i, \eta_j\right) = \mathrm{Cov}\left(\eta_{t+i}, \eta_{t+j}\right)$ for all t, i, and j,

where E and Cov stand for the expectation and covariance, respectively.

The parameters $\phi_1, \ldots, \phi_p$ are the unknown parameters of the AR(p) process and must lie in a certain region denoted by C_p for the process to be stationary, see Wise (1956) for more information about the region C_p. For example, for an AR(1) process to be stationary, we must have $|\phi_1| < 1$ and for an AR(2) process we must have $|\phi_2| < 1$ and $|\phi_1| < (1 - \phi_2)$.

The covariance matrix Σ has the form

$$\Sigma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix},$$

where γ_j=Cov $\left(\eta_t, \eta_{t+j}\right)$, $j = 0, \ldots, p-1$. To find the relation between $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$, one can make use of the Yule-Walker equations (Yule, 1927; Walker, 1931) which can be represented in the following matrix notation

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

and the equation

$$\gamma_0 = \sum_{k=1}^{p} \phi_k \gamma_k + \sigma_V^2. \tag{2.3}$$

The joint distribution of $\boldsymbol{\eta} = \{\eta_1, \ldots, \eta_n\}$ is given by

$$f\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_V^2\right) = f_{\eta_{1:p}}\left(\eta_1, \ldots, \eta_p|\boldsymbol{\phi}, \sigma_V^2\right) \times \prod_{t=p+1}^{n} \varphi_N\left(\eta_t; \mu_t, \sigma_V^2\right), \tag{2.4}$$

where $f_{\eta_{1:p}}$ is the joint density of $\eta_1, ..., \eta_p$, and $\varphi_N(\eta_t; \mu_t, \sigma_V^2)$ is the density of η_t given $\eta_1, \ldots, \eta_{t-1}$, which is the Gaussian distribution with argument η_t, mean μ_t and variance σ_V^2. The mean μ_t is given by

$$\mu_t = \phi_1 \eta_{t-1} + \phi_2 \eta_{t-2} + \cdots + \phi_p \eta_{t-p}.$$

2.2.1 Prior Distributions

In general, the prior distribution reflects the prior knowledge or information about the parameter of interest before collecting the data. If there is no such knowledge, then weakly-informative priors need to be assigned to the model parameters. In this section, we will use weakly-informative priors that have been proposed or discussed by other researchers.

1. Prior for $\boldsymbol{\beta}$: The regression parameters will be assigned independent normal priors with zero mean and standard deviation of 100, i.e., β_k are $i.i.d$ $N(\mu_\beta = 0, \sigma_\beta = 100)$, $k = 0, \ldots, K$.

2. Prior for $\boldsymbol{\phi}$: If the stationary condition is to be satisfied, the parameters $\{\phi_1, \ldots, \phi_p\}$ are constrained to lie in the region C_p. The form of this region is easy to identify for $p \leq 2$, but becomes complicated for $p > 2$. For this reason, we work on transformed variables where the domain is the entire real line. We will first reparametrize $\{\phi_1, \ldots, \phi_p\}$ in terms of the partial autocorrelations $\{r_1, \ldots, r_p\}$ as given below. The partial autocorrelation is the linear dependence of a variable with itself at two time points after removing any linear dependence on other time points. This transformation has been discussed in Jones (1987).

$$z_i^{(k)} = z_i^{(k-1)} - r_k z_{k-i}^{(k-1)}, \quad i = 1, ..., k-1, \tag{2.5}$$

with $z_k^{(k)} = r_k$ and $\phi_k = z_k^{(p)}$, for $k = 1, ..., p$. The Jacobian of this transformation is given by

$$J(r) \propto \prod_{k=2}^{p} (1 - r_k)^{[k/2]} (1 + r_k)^{[(k-1)/2]},$$

for $p \geq 2$ and one when $p = 1$. The symbol $[k]$ denotes the integer part of k.

We follow Jones (1987) and set priors on $\{r_1, \ldots, r_p\}$ by letting $r_k = 2B_k - 1$, where $B_k, k = 1, \ldots, p$ are independent beta random variables as follows. This prior will lead to a uniform prior on $\boldsymbol{\phi}$ on the C_p region.

$$B_k \sim \text{Beta}\left(\left[\frac{1}{2}(k+1)\right], \left[\frac{1}{2}k\right] + 1\right). \tag{2.6}$$

The condition that $\boldsymbol{\phi} \in C_p$ now becomes $|\mathbf{r}| < 1$. We then apply a second transformation from r to r^* which is defined by.

$$r_k^* = \log\left(\frac{1+r_k}{1-r_k}\right), \quad k = 1, ..., p.$$

The condition now becomes $(-\infty < r^* < \infty)$. Using the prior on r_k as in Eq. (2.6) and by applying the change of variable technique, the prior on r_k^* will be,

$$\pi_{r_k^*}(r_k^*) \propto \left(\frac{e^{r_k^*}}{e^{r_k^*}+1}\right)^{[\frac{1}{2}(k+1)]-1} \left(\frac{1}{e^{r_k^*}+1}\right)^{[\frac{1}{2}k]} \left(\frac{e^{r_k^*}}{\left(e^{r_k^*}+1\right)^2}\right), \quad r_k^* \in \mathbb{R}. \tag{2.7}$$

3. Prior on σ_V: We follow Gelman (2006) and use a uniform prior on σ_V over its entire domain. This prior is preferred over the inverse gamma prior in which case the results is affected by the choice of the inverse gamma parameters.

   Using the uniform prior on σ_V, the prior on σ_V^2 is

$$\pi_{\sigma_V^2}\left(\sigma_V^2\right) \propto \left(\sigma_V^2\right)^{-\frac{1}{2}} \quad \sigma_V^2 \in \mathbb{R}^+.$$

We will do all the random generation in the space of r^* and σ_V^2 and then transform back to ϕ and σ_V.

The posterior distribution in Eq. (2.2) does not seem to belong to a well-known distribution family. Therefore, it is necessary to turn to simulation-based methods to obtain posterior summaries and perform Bayesian inference. Below, we describe how to sample from this posterior.

2.3 The Single Move Method

In this section, we propose an accept-reject method to sample the latent variables $\boldsymbol{\eta}$ one at a time from their full conditional posterior. Sampling the latent process has some advantages.

It allows us to avoid evaluating the high dimensional integral involved in the likelihood function, which is hard to evaluate using numerical methods. It also makes the model selection and model adequacy checking feasible as both of these tasks involve the evaluation of the likelihood function. The regression parameters $\boldsymbol{\beta}$ and $\mathbf{r}^*$ will be sampled using the M-H algorithm (see Sections 2.3.2 and 2.3.3), and σ_V^2 will be sampled from an inverse Gamma distribution (see Section 2.3.4).

Before we continue, we will rewrite the joint distribution of $\boldsymbol{\eta}$, $f\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_V^2\right)$, in a convenient way for the SM method. First, please note that the joint distribution of $\eta_1, \ldots, \eta_p$ can be written as a product of conditional normal distributions. Second, it has been shown that we can factor σ_V^2 out from Σ, i.e., we can write $\Sigma = \sigma_V^2 \Sigma^*$, where Σ^* depends only on $\boldsymbol{\phi}$, see for example, Brockwell and Davis (1991) page 249. Using these facts, we can write $f_{\boldsymbol{\eta}}\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_v^2\right)$ as follows:

$$f_{\boldsymbol{\eta}}\left(\boldsymbol{\eta}|\boldsymbol{\phi}, \sigma_v^2\right) = \varphi_N\left(\eta_1; \mu_1^*, \sigma_V^2 R_1\right) \left\{ \prod_{t=2}^{n} \varphi_N\left(\eta_t; \mu_t^*, \sigma_V^2 R_t\right) \right\}, \tag{2.8}$$

where, for $t > 1$, $\mu_t^* = \mathrm{E}\left(\eta_t|\eta_1, \ldots, \eta_{t-1}\right)$, and $\sigma_V^2 R_t = \mathrm{Var}\left(\eta_t|\eta_1, \ldots, \eta_{t-1}\right)$. For $t = 1$, $\mu_1^* = \mathrm{E}\left(\eta_1\right)$ and $\sigma_V^2 R_1 = \mathrm{Var}\left(\eta_1\right)$. Here, R_t are functions of $\boldsymbol{\phi}$ alone.

In Eq. (2.8), $\mu_t^* = \sum_{j=1}^{p} \phi_j \eta_{t-j}$ and $R_t = 1$ for $t \geq p+1$. This can be seen from Eq. (1.3). For $t = 1$, we have $\mu_1^* = 0$ and $R_1 = \left\{ 1 - \sum_{j=1}^{p} \rho_j \phi_j \right\}^{-1}$, where $\rho_j = \dfrac{\gamma_j}{\gamma_0}$. R_1 can be obtained from Eq. (2.3).

For the remaining t, $(1 < t < p+1)$, μ_t^* and R_t can be obtained using the properties of the conditional mean and variance of multivariate normal distribution as follows: let $\boldsymbol{\eta}_{t-1} = \{\eta_1, \ldots, \eta_{t-1}\}$, for $t = 2, \ldots, p$, with the joint distribution being multivariate normal. The mean of this multivariate normal is zero and the covariance matrix is Σ_{t-1}, is given by

$$\Sigma_{t-1} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{t-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{t-3} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{t-2} & \gamma_{t-3} & \cdots & \gamma_0 \end{bmatrix}.$$

Also, let $\Sigma_{t,(t-1)}$ be the covariance vector between η_t and $\boldsymbol{\eta}_{t-1}$ which is given by

$$\Sigma_{t,(t-1)} = (\gamma_{t-1}, \gamma_{t-2}, \ldots, \gamma_1)\,.$$

Given that the marginal mean of η_t is zero, the conditional mean of $\eta_t|\boldsymbol{\eta}_{t-1}$ is given by

$$\mu_t^* = \Sigma_{t,(t-1)}\Sigma_{t-1}^{-1}\boldsymbol{\eta}_{t-1},$$

and the conditional variance is given by

$$\sigma_V^2 R_t = \gamma_0 - \Sigma_{t,(t-1)}\Sigma_{t-1}^{-1}\Sigma_{t,(t-1)}^T.$$

For example, to find the conditional mean and variance for $\eta_2|\eta_1$, we have the conditional mean as

$$\begin{aligned}\mu_2^* &= \gamma_1\gamma_0^{-1}\eta_1 \\ &= \rho_1\eta_1,\end{aligned}$$

and the conditional variance as

$$\begin{aligned}\sigma_V^2 R_2 &= \gamma_0 - \gamma_1\gamma_0^{-1}\gamma_1 \\ &= \gamma_0(1-\rho_1^2),\end{aligned}$$

which means $R_2 = R_1\left\{1-\rho_1^2\right\}$. We can continue this way to find μ_t^* and R_t for all $2 \leq t \leq p$.

2.3.1 Sampling the Latent Process

The most time consuming part is the sampling of the latent variables, as it has to be carried out n times. In the $AR\,(p)$ process, the distribution of η_t depends only on the nearest p variables; as a result, the conditional posterior density of η_t can be written as

$$\pi\left(\eta_t \mid \mathbf{y}, \eta_{(-t)}, \boldsymbol{\theta}\right) \propto \pi\left(\eta_t \mid y_t, \eta_{(-t)}, \boldsymbol{\theta}\right) \tag{2.9}$$

$$\propto g\left(y_t|\eta_t, \boldsymbol{\beta}\right) \prod_{j=t}^{t+min(p,n-t)} \varphi_N\left(\eta_j; \mu_j^*, \sigma_V^2 R_j\right), \tag{2.10}$$

where $\eta_{(-t)} = \{\eta_1, \ldots, \eta_{t-1}, \eta_{t+1}, \ldots, \eta_n\}$. In the full conditional posterior of η_t we are conditioning on the current observation y_t not the whole vector of observations. This is true due to the conditional independence assumption of y_t given η_t.

The formula above can be written in the following form:

$$\pi\left(\eta_t \mid y_t, \eta_{(-t)}, \boldsymbol{\theta}\right) \propto g\left(y_t|\eta_t, \boldsymbol{\beta}\right) \varphi_N\left(\eta_t; \frac{\mu_t}{l_t}, \frac{\sigma_V^2}{l_t}\right). \tag{2.11}$$

where μ_t and l_t are determined by $\pi(\eta_t|\eta_{(-t)})$. In an AR(2) process, for example, we have

$$\mu_t = \begin{cases} \phi_1\eta_2 + \phi_2\eta_3 & t = 1 \\ \phi_1\left(\eta_1 + \eta_3\right) + \phi_2\left(\eta_4 - \phi_1\eta_3\right) & t = 2 \\ \phi_1\left(1 - \phi_2\right)\left(\eta_{t-1} + \eta_{t+1}\right) + \phi_2\left(\eta_{t-2} + \eta_{t+2}\right) & t = 3, \ldots, n-2 \\ \phi_1\left(\eta_n + \eta_{n-2}\right) + \phi_2\left(\eta_{n-3} - \phi_1\eta_{n-2}\right) & t = n-1 \\ \phi_1\eta_{n-1} + \phi_2\eta_{n-2} & t = n \end{cases},$$

and

$$l_t = \begin{cases} 1 & t = 1, n \\ 1 + \phi_1^2 & t = 2, n-1 \\ 1 + \phi_1^2 + \phi_2^2 & t = 3, \ldots, n-2 \end{cases}.$$

In AR(1), we simply set $\phi_2 = 0$ in the above to get the full conditionals.

We now propose the rejection method to sample from $\pi\left(\eta_t \mid y_t, \eta_{(-t)}, \boldsymbol{\theta}\right)$. In this rejection method, we will use the tangent line of the function $\log g\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)$ to propose an envelop function and perform the accept-reject method. We can show that the function $\log g\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)$ is concave function in η_t which means that it can be bounded from above by its tangent line. Let $\log g\left(y_t \mid \eta_t, \boldsymbol{\beta}\right) = C + \log g^*\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)$, where C is constant and $\log g^*\left(y_t \mid \eta_t, \boldsymbol{\beta}\right) = -\exp\left(\boldsymbol{\beta}^T X_t + \eta_t\right) + y_t\eta_t$. The tangent line of g^* at $\eta_t = m_0$, where m_0 is the mode of the posterior in Eq. (2.11), is given by

$$-\exp\left(\boldsymbol{\beta}^T X_t + m_0\right) + y_t m_0 + \left\{-\exp\left(\boldsymbol{\beta}^T X_t + m_0\right) + y_t\right\}\left(\eta_t - m_0\right).$$

Rearranging the terms in the above equation in terms of η_t, it can be expressed as

$$\exp\left(\boldsymbol{\beta}^T X_t + m_0\right)\left(m_0 - 1\right) + \eta_t\left\{y_t - \exp\left(\boldsymbol{\beta}^T X_t + m_0\right)\right\}.$$

Now we have

$$\log g^*\left(y_t \mid \eta_t, \boldsymbol{\beta}\right) \leq \exp\left(\boldsymbol{\beta}^T X_t + m_0\right)(m_0 - 1) + \eta_t\left\{y_t - \exp\left(\boldsymbol{\beta}^T X_t + m_0\right)\right\}.$$

Hence:

$$\begin{aligned} g^*\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)\varphi_N\left(\eta_t; \frac{\mu_t}{l_t}, \frac{\sigma_V^2}{l_t}\right) \leq \quad & \exp\left\{\exp\left(\boldsymbol{\beta}^T X_t + m_0\right)(m_0 - 1)\right\} \times \\ & \exp\left\{\eta_t\left\{y_t - \exp\left(\boldsymbol{\beta}^T X_t + m_0\right)\right\}\right\} \times \\ & \varphi_N\left(\eta_t; \frac{\mu_t}{l_t}, \frac{\sigma_V^2}{l_t}\right). \end{aligned}$$

The two terms that include η_t on the right-hand side can be combined and shown to be proportional to the normal distribution $\varphi_N\left(\eta_t; \mu_t^0, \frac{\sigma_V^2}{l_t}\right)$ where

$$\mu_t^0 = \left\{\frac{\mu_t}{l_t} + \frac{\sigma_V^2}{l_t}\left(y_t - \exp\left(\boldsymbol{\beta}^T X_t + m_0\right)\right)\right\}.$$

With the above result, the accept-reject method to sample η_t from $\pi\left(\eta_t \mid y_t, \eta_{(-t)}, \boldsymbol{\theta}\right)$ can now be implemented.

Step 1: Sample η_t from $\varphi_N\left(\eta_t; \mu_t^0, \frac{\sigma_V^2}{l_t}\right)$, and u from the uniform distribution, U(0,1).

Step 2: If $u < \dfrac{g^*\left(y_t \mid \eta_t, \boldsymbol{\beta}\right)\varphi_N\left(\eta_t; \frac{\mu_t}{l_t}, \frac{\sigma_V^2}{l_t}\right)}{\exp\left\{\exp\left(\boldsymbol{\beta}^T X_t + m_0\right)(m_0 - 1)\right\}\exp\left\{((\mu_t^0)^2 - \mu_t^2)/(2\sigma_V^2 l_t)\right\}\varphi_N\left(\eta_t; \mu_t^0, \frac{\sigma_V^2}{l_t}\right)}$ accept the sampled value η_t. Otherwise, return to Step 1 to draw a new sample.

Please note that the mode m_0, was obtained using the Newton-Raphson method. In the simulation studies, we notice that the Newton-Raphson method was very sensitive to the initial guess, especially if the observed count y_t is very large. This is because the derivative is infinity for initial guess equal to zero, but larger initial guess solves the problem. In such a case, we observed that an initial guess close to $\log(y_t)$ works well.

2.3.2 Sampling the Poisson Regression Parameters $\boldsymbol{\beta}$

The regression parameters, $\boldsymbol{\beta}$, can be sampled from the posterior distribution that is proportional to

$$\pi\left(\boldsymbol{\beta} \mid \boldsymbol{\eta}, \mathbf{y}\right) \propto \prod_{t=1}^{n} g\left(y_t | \eta_t, \boldsymbol{\beta}\right) \prod_{i=0}^{k} \varphi_N\left(\beta_i; \mu_\beta, \sigma_\beta^2\right). \tag{2.12}$$

Sampling from this posterior can be accomplished using the M-H algorithm with a normal proposal. The mean and variance of this normal proposal were found by optimizing the log posterior in Eq. (2.12), where the mean was taken to be the mode of the log posterior and the variance to be the inverse of the Hessian matrix evaluated at the mode. Again, the mode was obtained using the Newton-Raphson method. We sampled the $\boldsymbol{\beta}$ parameters independently using a sequence of the M-H method, however, if there is a high dependence between elements of $\boldsymbol{\beta}$, then we can sample them as a block, or at least partition them into low dimensional blocks and then sample the blocks by a sequence of the M-H steps.

2.3.3 Sampling $\left\{r_1^*, \ldots, r_p^*\right\}$

The full conditional posterior of $\left\{r_1^*, \ldots, r_p^*\right\}$ is given by

$$\begin{aligned}
\pi\left(r_1^*, \ldots, r_p^* | \boldsymbol{\beta}, \sigma_V^2, \boldsymbol{\eta}, \mathbf{y}\right) &= f_{\boldsymbol{\eta}}\left(\eta_1, \ldots, \eta_t | \boldsymbol{\phi}, \sigma_V^2\right) \prod_{k=1}^{p} \pi_{r_k^*}\left(r_k^*\right) \\
&\propto \left\{\prod_{t=1}^{p} R_t\right\}^{-1} \exp\left\{-\frac{\sum_{t=1}^{n}\left(\eta_t - \mu_t^*\right)^2 / 2R_t}{\sigma_V^2}\right\} \\
&\times \prod_{k=1}^{p} \pi_{r_k^*}\left(r_k^*\right).
\end{aligned}$$

Sampling from this posterior can be accomplished using the M-H algorithm with independent normal proposals. The means of these normal proposals are the previously sampled values, and the variances are chosen such that the acceptance rate of the M-H is between 0.5 to 0.6.

2.3.4 Sampling σ_V^2

The full conditional posterior of σ_V^2 is given by

$$\begin{aligned}
\pi\left(\sigma_V^2|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\eta},\mathbf{y}\right) &= \pi\left(\sigma_V^2|\boldsymbol{\phi},\boldsymbol{\eta}\right) \\
&= f_{\boldsymbol{\eta}}\left(\eta_1,\ldots,\eta_t|\boldsymbol{\phi},\sigma_V^2\right)\pi_{\sigma_V^2}\left(\sigma_V^2\right) \\
&\propto \left(\sigma_V^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{\sum_{t=1}^{n}\left(\eta_t-\mu_t^*\right)^2/2R_t}{\sigma_V^2}\right\} \\
&\times\pi_{\sigma_V^2}\left(\sigma_V^2\right) \\
&= \left(\sigma_V^2\right)^{-\frac{n-1}{2}+1}\exp\left\{-\frac{\sum_{t=1}^{n}\left(\eta_t-\mu_t^*\right)^2/2R_t}{\sigma_V^2}\right\}.
\end{aligned}$$

Hence, σ_V^2 can be sampled from inverse gamma distributions, IG(α,γ), with parameters $\alpha=\frac{n-1}{2}$ and $\gamma=\sum_{t=1}^{n}\left(\eta_t-\mu_t^*\right)^2/2R_t$.

Chapter 3

The Particle Gibbs Sampler

3.1 Motivation

The goal of this chapter is to investigate and implement the Particle Gibbs Sampler (PGS) method and to compare it with the SM method discussed in Chapter 2. The PGS is a Particle Markov chain Monte Carlo (PMCMC) method developed by Andrieu et al. (2010) that combines MCMC methods and Sequential Monte Carlo (SMC) methods and takes the strength of the two components.

In the SM method, the latent variables $\boldsymbol{\eta}$ were sampled one by one using the accept-reject method. In order to improve the SM method, rather than sampling the latent process one by one, it may be helpful to sample the latent variables from their joint distribution as a block. In this chapter, we discuss the use of the PGS method to accomplish this. The hope is that sampling these variables as a whole will reduce the auto-correlation between posterior samples and, in turn, will lead to a more efficient method. The PGS method skillfully combines MCMC and SMC methods, where SMC is used to construct a high dimensional proposal to sample dependent variables, like the latent variables, inside the MCMC methods. SMC methods were first discovered by Gordon et al. (1993), and were used together with MCMC by Gilks and Berzuini (2001) and have been proved to be an effective combination between the computational advantages of SMC algorithms and the efficiency of the MCMC methods. While Gilks and Berzuini (2001) used an MCMC kernel to build a proposal distribution for SMC algorithm, Andrieu et al. (2010) pursued a totally different approach; they used SMC methods to build an efficient high dimensional proposal for the MCMC methods.

The naive use of the SMC methods for sampling the latent variables within the Gibbs sampler will not give a sample from the correct posterior distribution. In Andrieu et al. (2010), the PGS method is proposed in order to make the posterior be the target distribution of the sampling method. The key ingredient of the PGS method is the so-called conditional SMC, which resembles ordinary SMC with a pre-specified path, which we shall call the conditioned path. The conditioned path is ensured to survive all the sampling and resampling steps.

3.2 Sequential Monte Carlo Methods

Sequential Monte Carlo methods are a group of techniques combining importance sampling and resampling steps to sample recursively from a joint probability distribution, $\pi\left(\eta_1, \ldots, \eta_n|\boldsymbol{\theta}\right)$, which is known up to a normalizing constant. SMC methods employ a population of samples (particles) and a sequence of 'intermediate' probability distributions of increasing dimension, with the final distribution being the target distribution, to generate samples from the target. The intermediate distribution are denoted by $\left\{\pi_t\left(\eta_{1:t}|\boldsymbol{\theta}\right), t = 1, \ldots, n\right\}$, with $\eta_{1:t} = (\eta_1, \ldots, \eta_t)$ and $\pi_n\left(\eta_{1:n}|\boldsymbol{\theta}\right) = \pi(\eta_{1:n}|\boldsymbol{\theta})$. Each of these densities is assumed to be known up to a normalizing constant, that is, for $t = 1, \ldots, n$

$$\pi_t\left(\eta_{1:t}|\boldsymbol{\theta}\right) = \frac{p_t\left(\eta_{1:t}|\boldsymbol{\theta}\right)}{Z_t},$$

where Z_t is the normalizing constant and is given by

$$Z_t = \int \cdots \int p_t(\eta_{1:t}|\boldsymbol{\theta}) d\eta_1 \ldots d\eta_t.$$

The main idea of the SMC methods is as follows. To start, we need to specify an importance density $q_1\left(\eta_1|\boldsymbol{\theta}\right)$ to initialize the recursion at time $t = 1$ and a family of associated transition densities $q_t\left(\eta_t|\eta_{1:(t-1)}, \boldsymbol{\theta}\right)$. At time $t = 1$, use importance sampling to draw N particles, denote them by $\eta_1^{1:N} = \{\eta_1^1, \ldots \eta_1^N\}$, from the importance density $q_1\left(\eta_1|\boldsymbol{\theta}\right)$ and assign importance weights to them

$$w_1^k = \frac{p_1\left(\eta_1^k|\boldsymbol{\theta}\right)}{q_1\left(\eta_1^k|\boldsymbol{\theta}\right)}, k = 1, \ldots, N.$$

The weights $\{w_1^1, \ldots, w_1^N\}$ are unnormalized. Therefore, let $\{W_1^1, \ldots, W_1^N\}$ be their normalized counterpart. To obtain N particles approximately distributed according to $\pi_1(\eta_1|\boldsymbol{\theta})$, we resample N times from the available particles according to the normalized weights. Denote the resampled particles by $\{\tilde{\eta}_1^1, \ldots, \tilde{\eta}_1^N\}$. At subsequent time points ($t \geq 2$), we aim to sample from $\pi_t(\eta_{1:t}|\boldsymbol{\theta})$, which can be done by extending the path, $\tilde{\eta}_{1:(t-1)}^{1:N}$, obtained at time $t-1$, through sampling $\eta_t^{1:N}$ from $q_t\left(\eta_t^k|\tilde{\eta}_{1:(t-1)}^k, \boldsymbol{\theta}\right)$. Again, the normalized importance weights W_t^k need to be computed and a resampling step needs to be performed to produce samples, $\tilde{\eta}_{1:t}^{1:N}$, approximately distributed according to $\pi_t(\eta_{1:t}|\boldsymbol{\theta})$, i.e.,

$$\hat{\pi}_t^N(\eta_{1:t}|\boldsymbol{\theta}) = \sum_{k=1}^{N} W_t^k \delta_{\eta_{1:t}^k}(d\eta_{1:t}),$$

where $\delta_{\eta_{1:t}^k}(d\eta_{1:t})$ is the Dirac delta function.

The SMC algorithm described above is standard in the sense that the resampling step is performed at each time point t. However, in many applications of the SMC methods the resampling step is only performed if the estimator $\hat{\pi}_t^N(\eta_{1:t}|\boldsymbol{\theta})$ is poor. This can be assessed by looking at the variability in the normalized weights. One popular measure of variability is the so-called effective sample size (ESS) at time t which is given by

$$\text{ESS} = \left(\sum_{k=1}^{N} \left(W_t^k\right)^2\right)^{-1}.$$

The ESS takes values between one and N, with a value of N indicating equally weighted particles. The rule is to sample when ESS falls below a threshold value N^*. One choice of N^* is $\frac{N}{2}$. See Del Moral et al. (2006) for more details.

The resampling step can be interpreted as how the k^{th} particle at time t chooses its parent at time $t-1$, so the resampling step basically consists of selecting an index for the parent particle, which we will denote by A_{t-1}^k. From now on, we will use the notation $\eta_{1:t}^{A_t^k}$ to represent the resampled particle instead of $\tilde{\eta}_{1:t}^k$.

The pseudo-code of the above SMC algorithm is given below. Following the notations of Andrieu et al. (2010), we denote by $\mathcal{F}\left(.|W_t^{1:N}\right)$ the discrete probability mass function on the set $\{1, \ldots, N\}$ with associated probabilities $W_t^{1:N} = \{W_t^1, \ldots, W_t^N\}$. Whenever the index k is used, we mean for each of $k = 1, \ldots, N$.

Step 1: at time $t = 1$,

(a) sample $\eta_1^k \sim q_1(.|\boldsymbol{\theta})$,

(b) compute and normalize the weights

$$w_1\left(\eta_1^k\right) = \frac{p_1\left(\eta_1^k|\boldsymbol{\theta}\right)}{q_1\left(\eta_1^k|\boldsymbol{\theta}\right)}, \tag{3.1}$$

$$W_1^k = \frac{w_1\left(\eta_1^k\right)}{\sum_{k=1}^{N} w_1\left(\eta_1^k\right)}.$$

Step 2: at times $t = 2, \ldots, n$,

(a) calculate ESS$=\left(\sum_{k=1}^{N}\left(W_{t-1}^k\right)^2\right)^{-1}$,

(b) if ESS $< N^*$, sample $A_{t-1}^k \sim \mathcal{F}\left(.|W_{t-1}^{1:N}\right)$. Otherwise let $A_{t-1}^k = k$,

(c) sample $\eta_t^k \sim q_t\left(.|\eta_{1:(t-1)}^{A_{t-1}^k}, \boldsymbol{\theta}\right)$ and set $\eta_{1:t}^k = \left(\eta_{1:(t-1)}^{A_{t-1}^k}, \eta_t^k\right)$,

(d) compute and normalize the weights

$$w_t\left(\eta_{1:t}^k\right) = \frac{p_t\left(\eta_{1:t}^k|\boldsymbol{\theta}\right)}{p_{t-1}\left(\eta_{1:(t-1)}^{A_{t-1}^k}|\boldsymbol{\theta}\right) q_t\left(\eta_t^k|\eta_{1:(t-1)}^{A_{t-1}^k}, \boldsymbol{\theta}\right)}, \tag{3.2}$$

$$W_t^k = \frac{w_t\left(\eta_{1:t}^k\right)}{\sum_{k=1}^{N} w_t\left(\eta_{1:t}^k\right)}.$$

To keep track of the genealogy of particles, we introduce the notation of the ancestral lineage $B_{1:n}^k = \{B_1^k, \ldots, B_n^k\}$ of the k^{th} particle $\eta_{1:n}^k$. This ancestral lineage B_t^k represents the index associated with the ancestor particle of $\eta_{1:n}^k$ at time t. This helps in describing the path of the k^{th} particle, such that we can write

$$\eta_{1:n}^k = \left(\eta_1^{B_1^k}, \eta_2^{B_2^k}, \ldots, \eta_n^{B_n^k}\right).$$

For example, the ancestral lineage of the first particle in Figure 3.1 (red coloured) is $B_{1:5}^1 = \{3, 4, 2, 2, 1\}$. The ancestral lineage is defined using the following backward recursion

$$\begin{aligned} B_n^k &= k, \\ B_t^k &= A_t^{B_{t+1}^k}, \quad t = n-1, \ldots, 1. \end{aligned}$$

3.3 The Particle Gibbs Sampler

The PGS alternately sample from the conditional posteriors $\pi(\theta|\eta_{1:n}, y_{1:n})$ and then from $\pi(\eta_{1:n}|\theta, y_{1:n})$. Sampling from the former will be done in the same way as in SM, and is thus not discussed here. Sampling from the latter will be accomplished using SMC methods. However, naively replacing a sample from $\pi(\eta_{1:n}|\theta, y_{1:n})$ by a sample from the SMC approximation, $\hat{\pi}_t^N(\eta_{1:t}|\boldsymbol{\theta})$, does not constitute a valid MCMC method. To address this issue, a special type of SMC must be used, which is the conditional SMC (Andrieu et al., 2010). The conditional SMC update is similar to a standard SMC algorithm but is such that a pre-defined fixed particle $\eta_{1:n}^{B_{1:n}}$ with its ancestral lineage $B_{1:n}$ is ensured to survive all the resampling step, while the remaining particles are generated as usual. Because of this, the resampling step has to be performed at each time point in the conditional SMC, in contrast to the standard SMC where the resampling is only performed when the discrepancy between particles is large or – equivalently – when ESS is small. The conditional SMC algorithm can be outlined as follows.

Step 1: let $\eta_{1:n} = \left(\eta_1^{B_1}, \eta_2^{B_2}, \ldots, \eta_n^{B_n}\right)$ be the conditioned path associated with the ancestral lineage $B_{1:n}$,

Step 2: at time $t = 1$,

(a) for $k \neq B_1$, sample $\eta_1^k \sim q_1(.|\boldsymbol{\theta})$,

(b) compute $w_1\left(\eta_1^k\right)$ by using equation(3.1) and normalize the weights.

Step 3: at times $t = 2, \ldots, n$,

(a) for $k \neq B_t$, sample $A_{t-1}^k \sim \mathcal{F}\left(.|W_{t-1}^{1:N}\right)$,

(b) for $k \neq B_t$, sample $\eta_t^k \sim q_t\left(.|\eta_{1:(t-1)}^{A_{t-1}^k}, \boldsymbol{\theta}\right)$ and set $\eta_{1:t}^k = \left(\eta_{1:(t-1)}^{A_{t-1}^k}, \eta_t^k\right)$,

(c) compute the weights by using equation (3.2) and normalize the weights.

When conditional SMC methods are used within MCMC methods, only one particle will be selected at time $t = n$, and the selected particle will be used to sample $\boldsymbol{\theta}$ from their conditional posterior. Now, let us introduce the notation for any variable X at iteration i of the MCMC by $X(i)$. The particle Gibbs sampler can now be implemented as follows.

Step 1: initialize the PGS: $i = 0$, set $B_{1:n}(0)$ and $\boldsymbol{\theta}(0)$ and $\left\{\eta_1^{B_1(0)}(0), \ldots, \eta_n^{B_n(0)}(0)\right\}$ arbitrarily.

Step 2: for iteration $i \geq 1$,

1. run a conditional SMC algorithm targeting $\pi(\boldsymbol{\eta}|\mathbf{y}, \boldsymbol{\theta}(i-1))$ conditional on $\left\{\eta_1^{B_1(i-1)}(i-1), \ldots, \eta_n^{B_n(i-1)}(i-1)\right\}$ and $B_{1:n}(i-1)$,
2. sample one particle randomly according to the weights and call it $\left\{\eta_1^{B_1(i)}(i), \ldots, \eta_n^{B_n(i)}(i)\right\}$. Hence, $B_{1:n}(i)$ is also implicitly sampled,
3. sample $\boldsymbol{\theta}(i)$ conditional on $\left\{\eta_1^{B_1(i)}(i), \ldots, \eta_n^{B_n(i)}(i)\right\}$, as discussed in Sections 2.3.2, 2.3.3 and 2.3.4.

3.3.1 Illustrative Example

We feel that conditional SMC methods deserve more illustration. In the following example, we will visualize the conditional SMC process. Suppose that we have a sample of size $n = 5$ and we want to use $N = 4$ particles in conditional SMC. Furthermore suppose that the conditioned path is $\{\eta_1^3, \eta_2^4, \eta_3^2, \eta_4^2, \eta_5^1\}$, with ancestral lineage $B_{1:4} = \{3, 4, 2, 2, 1\}$. The following is an illustration of the first two time points of Figure 3.1. At time $t = 1$, we need to make sure that η_1^3 will survive, so we need to sample three particles only ($\{\eta_1^1, \eta_1^2, \eta_1^4\}$). Now, at time $t = 2$, we know that the 4^{th} particle η_2^4 is connected with η_1^3, so we need this connection to survive, thus we need to sample only three father particle indicators ($\{A_1^1, A_1^2, A_1^3\}$), which indicate which particle at time $t = 1$ will be used to sample η_2. Suppose that we have $\{A_1^1 = 3, A_1^2 = 4, A_1^3 = 1\}$. That means we will sample η_2^1 from $q(\eta_2|\eta_1^3)$, η_2^2 from $q(\eta_2|\eta_1^4)$ and η_2^3 from $q(\eta_2|\eta_1^1)$. η_2^4 remains the same as it is in the conditioned path. We continue in

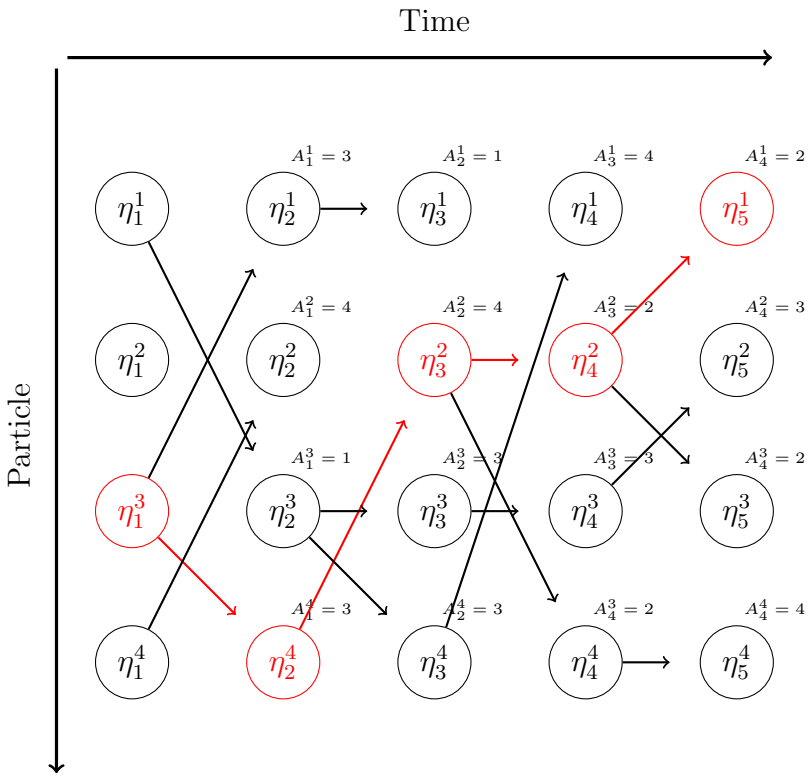

Figure 3.1: The structure of the conditional SMC

the same way until time point $t = n = 5$. At this time point, we need to select one particle according to the weights. Suppose the selected particle was particle number 2. We now need to figure out the ancestral lineage for the new particle. We will have $B_5 = 2$; the remaining B's can be found through the recursive relation $B_t = A_t^{B_{t+1}}$, $t = 4, \ldots, 1$, which means that $B_4 = A_4^2 = 3$, $B_3 = A_3^3 = 3$, $B_2 = A_2^3 = 3$, and $B_1 = A_1^3 = 1$, thus the newly selected path will be $\{\eta_1^1, \eta_2^3, \eta_3^3, \eta_4^3, \eta_5^2\}$; see Figure 3.2.

3.3.2 Resampling Step

There are many sampling procedures that can be found in the literature and are commonly used within the SMC methods. For comparisons between these methods, see Douc et al. (2005). In this thesis, we will consider using stratified resampling. Stratified resampling was

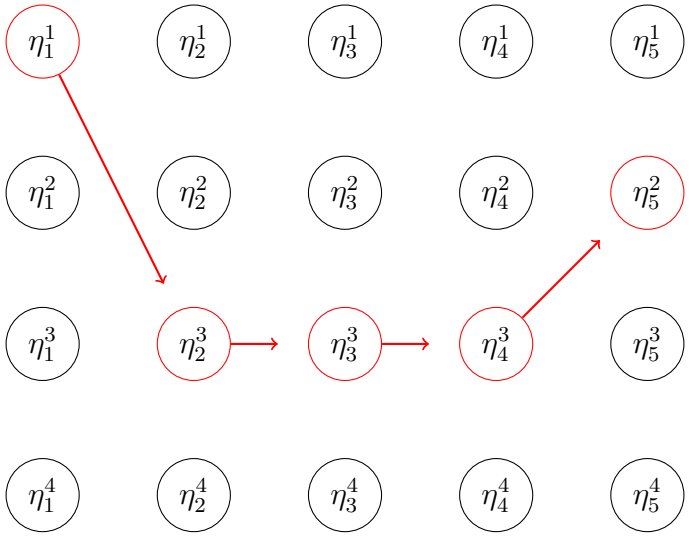

Figure 3.2: The selected path

originated in the context of survey sampling and was proved to be very useful technique in Monte Carlo computations (Liu, 2002).

In PGS, we want to use stratified resampling to sample $N-1$ father particles, with replacement, from the available N particles so that we make sure the conditioned particle remained the same. To do that, we first divide the interval (0,1) into $N-1$ sub-intervals of equal length, then we sample a uniform random number, $\{U_j, j = 1, \ldots, N-1\}$, from each sub-interval independently. Now, suppose for simplicity of presentation that at time t the conditioned particle was the first particle, i.e., $B_t = 1$. Then for $j = 2, \ldots, N$, let $A_{t-1}^j = k$, $k = 1, \ldots, N$ if

$$\left\{ \sum_{i=1}^{k-1} W_t^i \leq U_j \leq \sum_{i=1}^{k} W_t^i \right\}, \tag{3.3}$$

where $\sum_{i=1}^{0} W_t^i = 0$. Please note that it is guaranteed that there is at least one k that satisfies the condition in (3.3). This is because all U_j are in the interval (0,1) and the W_t^i, $i = 1, \ldots, N$ are also in the interval (0,1) and sum to one.

3.3.3 The Proposal Density

The choice of the proposal density plays an important role in SMC methods. A poorly chosen proposal function may lead to very inefficient sampling. A general rationale is to have a proposal density which retains the key features of the target distribution and does not deviate from the target distribution too much. The proposal density was chosen to be the t distribution with 15 degrees of freedom and mean and variance chosen according to the following.

We can show that, for all t, the conditional posterior of $\eta_t|\boldsymbol{\theta}, \mathbf{y}, \eta_{1:(t-1)}$ has the form

$$\pi(\eta_t|\boldsymbol{\theta}, \mathbf{y}, \eta_{1:(t-1)}) \propto \exp(-e^{\boldsymbol{\beta}^T\mathbf{X}_t+\eta_t} + y_t\eta_t) \times \varphi_N(\eta_t; \mu_t, \sigma_t^2). \tag{3.4}$$

The exponent in the exponential function above can be approximated using a Taylor series expansion about m_t, the mode of the posterior in eq.(3.4). Thus, we will have

$$\begin{aligned}
&-e^{\boldsymbol{\beta}^T\mathbf{X}_t+\eta_t} + y_t\eta_t \approx -e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t} + y_t m_t + \\
&(-e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t} + y_t)(\eta_t - m_t) - \frac{1}{2}e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t}(\eta_t - m_t)^2 \\
&= (-e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t} + y_t)\eta_t - \frac{1}{2}e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t}(\eta_t^2 - 2m_t\eta_t) + C,
\end{aligned} \tag{3.5}$$

where C is the part that does not depend on η_t, and which can therefore be ignored, as it will be combined with the normalizing constant of the density. After ignoring C in equation (3.5) above, the remaining part can be written as

$$-\frac{1}{2}\eta_t^2 e^{\boldsymbol{\beta}^T\mathbf{X}_t+m_t} + ((m_t - 1)e^{\boldsymbol{\beta}^T\mathbf{X}_t+mu_t} + y_t)\eta_t. \tag{3.6}$$

Combining the normal density in (3.4) and equation (3.6), we can show that the conditional posterior $\pi(\eta_t|\boldsymbol{\theta}, \mathbf{y}, \eta_{1:t-1})$ can be approximated by the following normal density

$$\pi(\eta_t|\boldsymbol{\theta}, \mathbf{y}, \eta_{1:t-1}) \approx \varphi_N\left(\eta_t; \mu_t^1, \sigma_t^{*^2}\right), \tag{3.7}$$

where

$$\mu_t^1 = \frac{\mu_t + \sigma_t^2 \left((m_t - 1) e^{\boldsymbol{\beta}^T \mathbf{X}_{t+m_t}} + y_t \right)}{\sigma_t^2 e^{\boldsymbol{\beta}^T \mathbf{X}_{t+m_t}} + 1},$$

and

$$\sigma_t^{*^2} = \frac{\sigma_t^2}{\sigma_t^2 e^{\boldsymbol{\beta}^T \mathbf{X}_{t+m_t}} + 1}.$$

The normal density in equation(3.7) provides a good proposal when the observed count y_t is small. However, we noticed that the target density $\pi(\eta_t|\boldsymbol{\theta}, \mathbf{y}, \eta_{1:(t-1)})$ is skewed when the observed count y_t is large, and thus this proposal may not cover both tails of the target. For this reason, we chose the t distribution to be the proposal density. The mean and variance of the t distribution were taken to be the same as those of the normal distribution in Eq. (3.7). Figure 3.3 shows how close are the target and proposal for one η_t. The corresponding observed count for this η_t was relatively large ($y_t = 156$). This plot is produced assuming $AR(2)$ with moderate correlation structure and large value of σ_V, as defined in Section 3.4.

3.3.4 The Number of Particles

Although Andrieu et al. (2010) showed theoretically that PGS will converge for any number of particles $N \geq 2$, we found using simulation that a small number of particles may lead to a computationally inefficient algorithm as a very large number of iterations in PGS is needed for the algorithm to converge. One reason for this inefficiency issue is the well known degeneracy issue in SMC methods. Andrieu et al. (2010) mentioned that PGS is less affected by the degeneracy issue than regular SMC methods. However, with a small number of particles, we may re-select the conditioned path or at least paths that have many particles in common with the conditioned path over many times. This means that we will use the same, or a large part, of the previously selected set $\boldsymbol{\eta}$'s in many successive iterations of PGS, which in turn will lead to slow convergence and additional required iterations in PGS. On the other hand, a large number of particles leads to heavy computational burden with little gain in the final results.

It is important to choose a proper number of particles. The way how we determine the

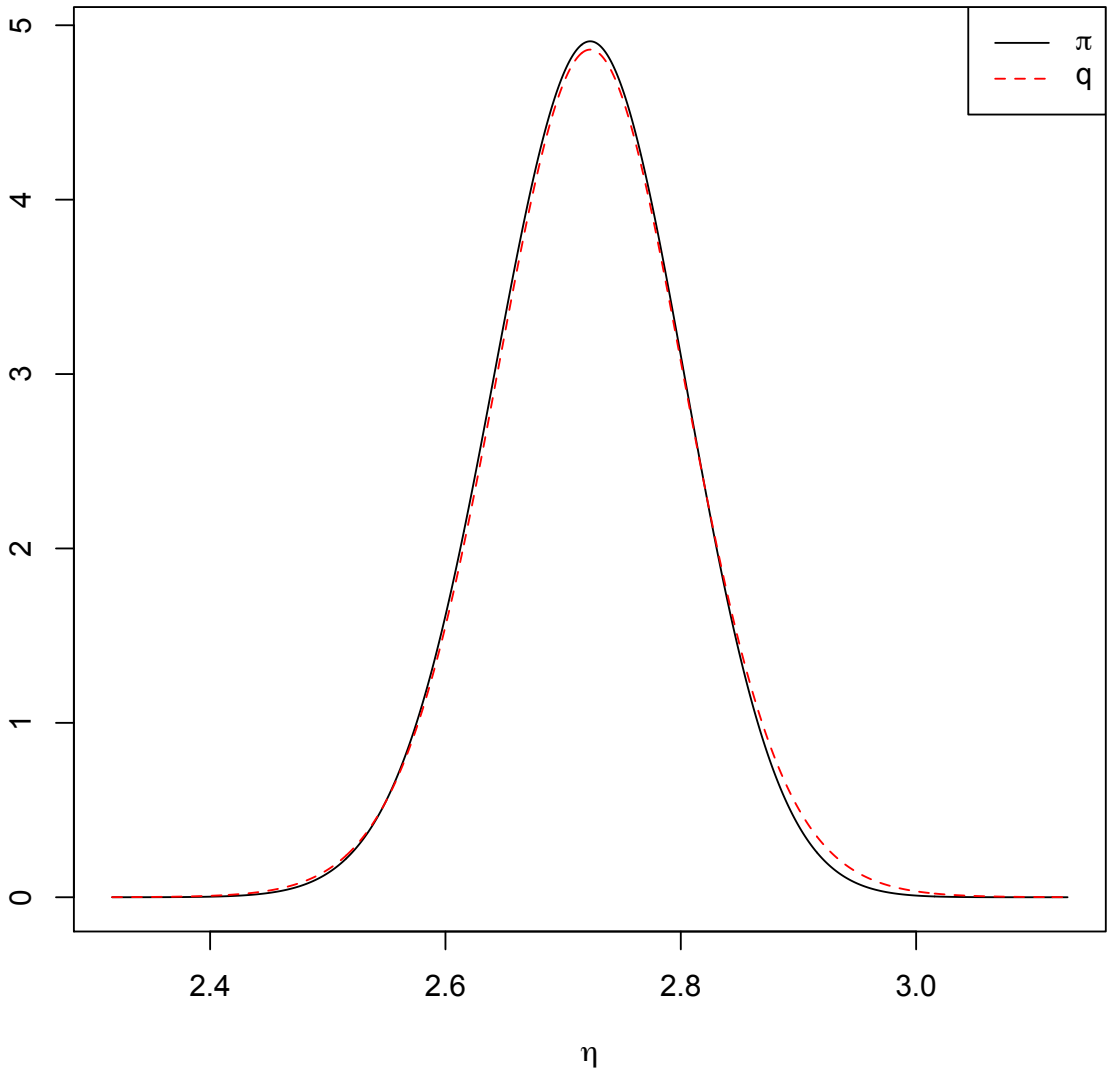

Figure 3.3: The true posterior $\pi(\eta_t|\eta_{1:(t-1)}, y_t, \boldsymbol{\theta})$ and the proposal density $q(\eta_t|\eta_{1:(t-1)}, y_t, \boldsymbol{\theta})$

number of particles is based on the mixing property of PGS, that is, how often the newly selected particle is *independent* from the conditioned path. By independent we mean, we do not select exactly the same particle. In the same spirit as M-H algorithm, we can look at the probability of selecting a new particle, given the previously selected particle. Figure 3.4 shows the probability (in percentage terms) of selecting a new particle when the sample size $n = 365$ for different values of N. It can be seen that the probability increases rapidly until $N \approx 800$ and the increase is slower after that; thus we use $N = 800$ when the sample size is $n = 365$.

3.4 Simulation Study Design and Results

The goal of this simulation study is to compare the PGS with the SM method discussed in Chapter 2. We will examine another simulation study in Chapter 4. To avoid repeating the simulation study design again in Chapter 4, we describe it here.

Although all methods have been discussed for any order, p, of the $AR(p)$ latent process and for any number of covariates, in the simulation study, we will focus on a situation similar to the real data example discussed in Chapter 5. We consider two orders of the autoregressive latent process, which are $p = 2$ and $p = 3$, with two covariates, one binary, and the other continuous. From the simulation study results we noticed that comparisons between the proposed methods were robust to the order of the latent process and lead to similar conclusions for $p = 2$ and $p = 3$, therefore we did not discuss a larger value of p.

For each choice of p, we discuss three different scenarios of the correlation structure in the latent process, that is, weak, moderate, and strong serial autocorrelation. We use two values of σ_V for each choice of p. For the regression parameters, we will have one set of values that are selected to be very close to the values found in the real data example. We will also study three sample sizes, $n = 20$, $n = 365$, and $n = 1095$. In total, we have $3 \times 2 \times 3$ different scenarios for each choice of p. For each scenario, we generate 200 data sets from the Poisson parameter driven model to perform the MC simulation study.

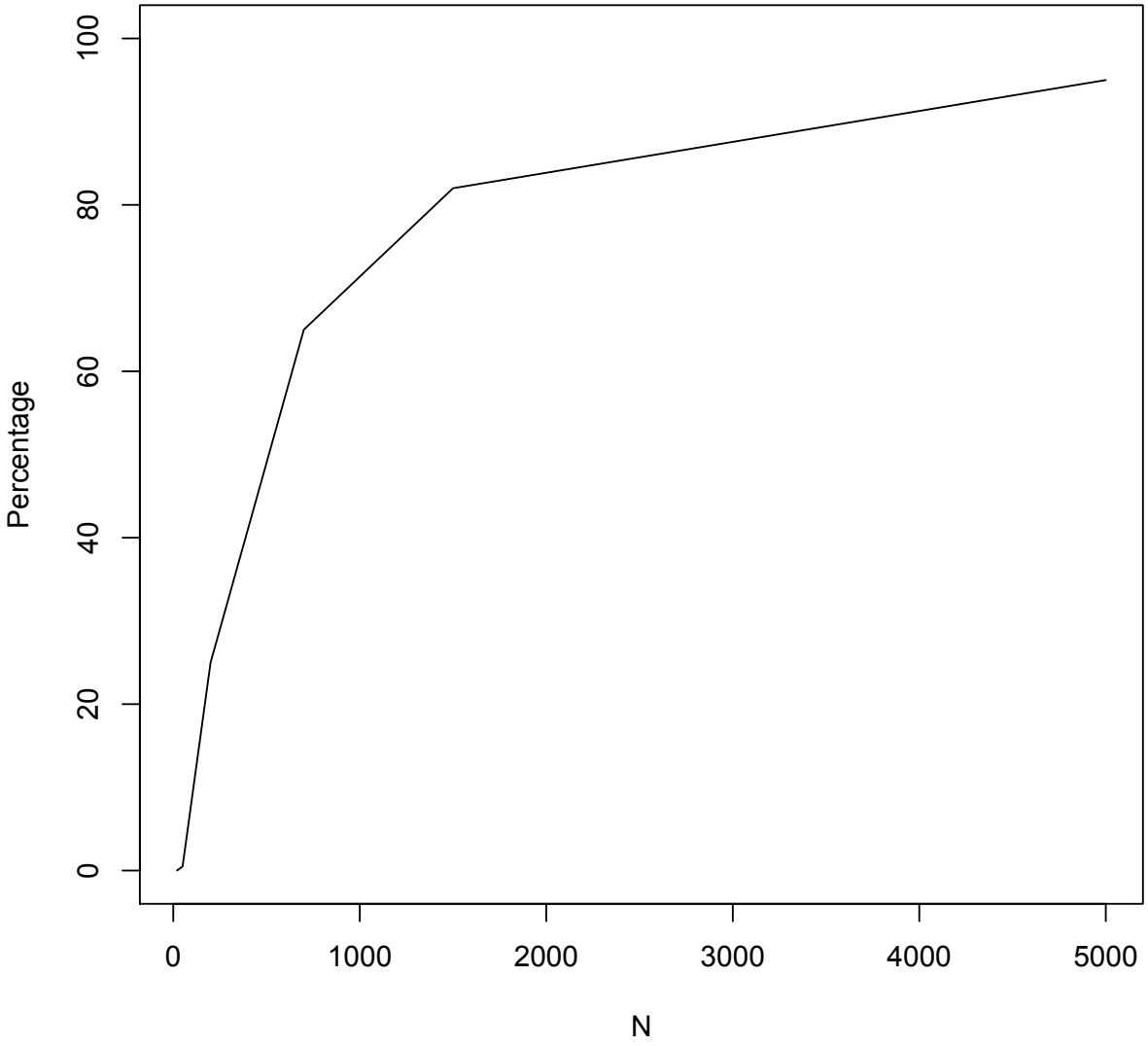

Figure 3.4: Percentage of times of selecting an independent particle

The simulation studies are performed using a cluster with the following specifications: two Intel Xeon CPU X5650 at 2.67 GHz, 24 GB of memory, and RAIDed hard drives with 2 TB storage. All of the code is written in the C language.

The true regression parameters were the same for both $AR(2)$ and $AR(3)$ processes, with values are $\boldsymbol{\beta}$=$\{1.5, -0.1, 0.4\}$. For the $AR(2)$ process, the two values of σ_V are 0.2 and 0.4. The vector has one large value and one small value, with the larger σ_V leading to larger serial correlations for Y_t (Ng et al., 2011). The three sets of (ϕ_1, ϕ_2) that correspond to the three correlation structures are as below.

1. $\phi_1 = 0.19$, $\phi_2 = 0.06$, this corresponds to the weak latent serial correlation $\rho_1 = 0.2$, $\rho_2 = 0.1$, where ρ_j is the serial correlation between η_t and η_{t+j}.

2. $\phi_1 = 0.40$, $\phi_2 = 0.20$, this corresponds to the moderate latent serial correlation $\rho_1 = 0.5$, $\rho_2 = 0.4$.

3. $\phi_1 = 0.85$, $\phi_2 = 0.1$, this corresponds to the strong latent serial correlation $\rho_1 = 0.95$, $\rho_2 = 0.90$.

For the $AR(3)$ latent process, the two values of σ_V are 0.3 and 0.5 and the three sets of (ϕ_1, ϕ_2, ϕ_3) that correspond to the three correlation structures are:

1. $\phi_1 = 0.20$, $\phi_2 = 0.10$, $\phi_3 = 0.10$, this corresponds to the weak latent serial correlation $\rho_1 = 0.20$, $\rho_2 = 0.16$, $\rho_3 = 0.10$.

2. $\phi_1 = 0.40$, $\phi_2 = 0.20$, $\phi_3 = 0.10$, this corresponds to the moderate latent serial correlation $\rho_1 = 0.56$, $\rho_2 = 0.48$, $\rho_3 = 0.40$.

3. $\phi_1 = 0.88$, $\phi_2 = 0.34$, $\phi_3 = -0.28$, this corresponds to the strong latent serial correlation $\rho_1 = 0.95$, $\rho_2 = 0.91$, $\rho_3 = 0.85$

We generated the autocorrelation (ACF) plots for the posterior samples of the parameters using both the SM and the PGS methods. We observed that there is no noticeable difference between the two methods. That is, the PGS method did not speed up the mixing. Figure

3.5 shows the ACF plot for posterior samples using the SM and PGS methods with a sample of size 365, and $AR(2)$ latent process with moderate serial correlation and larger value of σ_V.

The proposed methods are compared through the bias and the root mean square error (RMSE), which are defined as follows. Let $\boldsymbol{\theta}_0$ be the true parameter value and $\hat{\boldsymbol{\theta}}^{(j)}$ and $var^{(j)}$ are the mean and the variance of the posterior samples obtained from the j^{th} dataset. The bias and RMSE can be defined respectively as follows

$$\text{bias} = \frac{\sum_{j=1}^{200} \left(\hat{\boldsymbol{\theta}}^{(j)} - \boldsymbol{\theta}_0 \right)}{200},$$

and

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{200} \left(var^{(j)} + \left(\hat{\boldsymbol{\theta}}^{(j)} - \boldsymbol{\theta}_0 \right)^2 \right)}{200}}.$$

The conclusion that can be made based on the $AR(3)$ latent process is similar to the one based on the $AR(2)$ latent process. Also, a similar conclusion can be obtained based on the two values of σ_V. Therefore, only results of the $AR(2)$ latent process with the larger value of σ_V will be shown in this chapter. The complete results can be found in Tables A.1 to A.6 in Appendix A. The results of the AR(2) latent process with the larger value of σ_V are shown in Figures 3.6 to 3.8.

A general observation is that both methods tend to produce higher RMSE for the regression parameters, $\boldsymbol{\beta}$, for the case with larger σ_V compared to the case with smaller σ_V. The opposite is true for the latent process parameters, $\boldsymbol{\phi}$ and σ_V, where the case with larger σ_V produces smaller RMSE compared to the case with smaller σ_V. This observation seems to disappear as the sample size becomes larger. It can be seen also that the bias and RMSE of the parameter estimates fall as the sample size increases.

Some observations for small sample size, $n = 20$, are as follows. The SM method tends to have smaller RMSE for $\boldsymbol{\beta}$ than the PGS under all correlation structures. This is also true for the bias, except for β_1 under weak and moderate correlation structure, where the PGS has smaller bias than the SM. The opposite is true for σ_V and $\boldsymbol{\phi}$, where the PGS has smaller

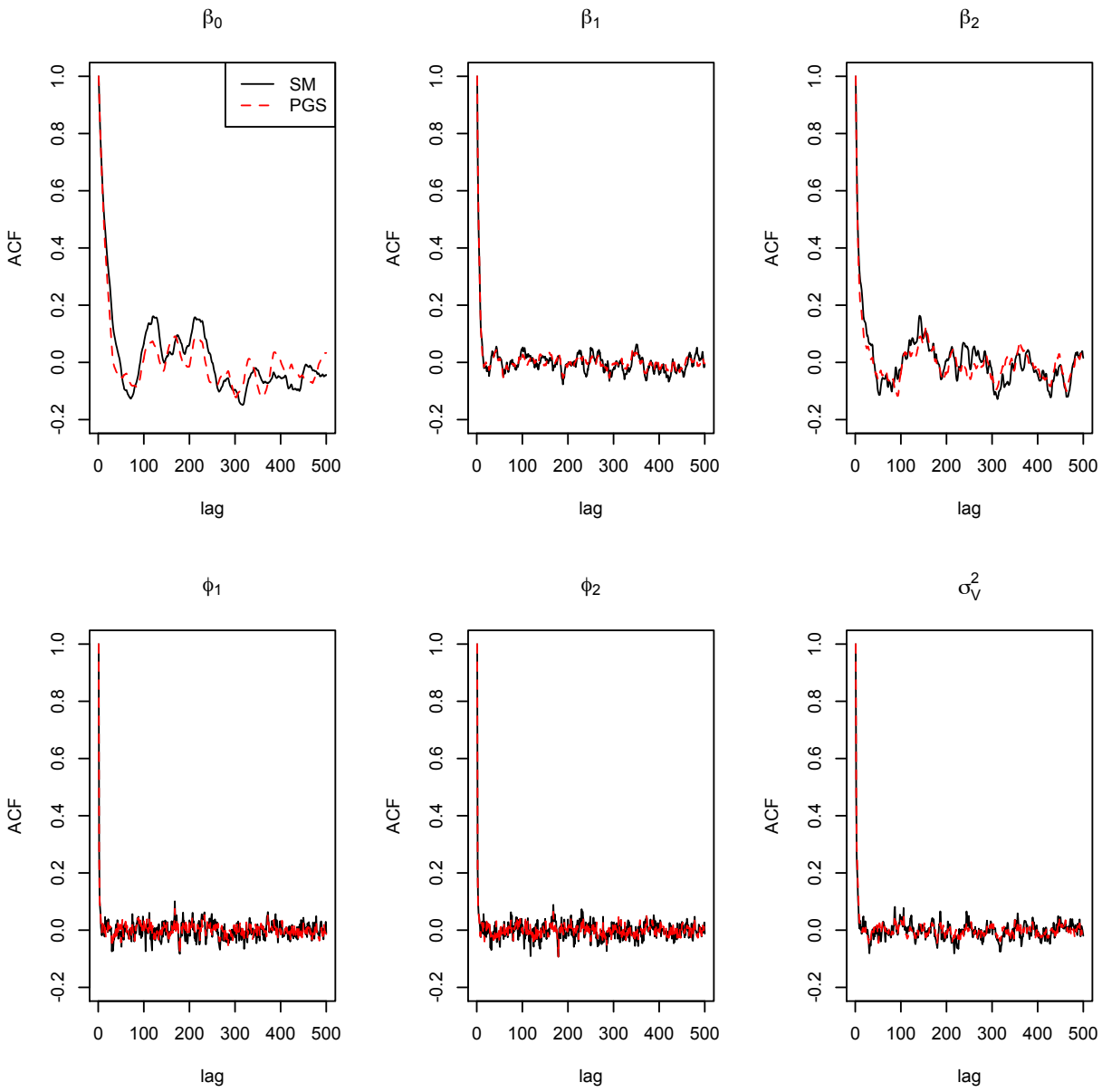

Figure 3.5: ACF plot for posterior samples using SM and PGS methods for $AR(2)$ latent process with moderate serial correlation and larger value of σ_V.

bias and RMSE than the SM. However, there are two exceptions; the first one is for σ_V with the strong serial correlation, where the SM has smaller bias and RMSE than the PGS, and the second is for ϕ_2 with weak serial correlation where the SM has smaller RMSE than the PGS.

As the sample size rises, the two methods tend to perform almost the same in terms of bias and RMSE for weak and moderate correlation structure. This is also true for the strong serial correlation structure; the exception is for β_0 where the PGS tends to have smaller bias and RMSE than the SM. In the plots below there might be no bars seen for the case of large sample size, and this is due to the fact that the bias and RMSE decrease to zero as the sample size increases.

In terms of the computational time, the SM is much faster than the PGS. With an Intel Xeon 2.67 GHz processor, the computational times in a C program are 15 seconds and around 100 minutes for SM and PGS, respectively, for 28000 MCMC iterations and a sample size of $n = 365$. In addition, the the computational time for the SM method increases linearly as the sample size increases, while it increases in higher scale for the PGS method. This is due to the fact that in the PGS method a larger number of particles are needed for larger sample sizes.

In general, we conclude that the SM method outperforms, or is at least as good as, the PGS in all cases in terms of the RMSE of regression parameters. The only exception is that the PGS outperforms the SM for β_0 only with strong serial correlation and large sample size. Thus, since, in the majority of applications, the main interest is in the regression parameters, excluding β_0, we recommend the use of the SM method because: (1) it provides similar results and is less time consuming than the PGS for large sample sizes, and (2) it gives smaller RMSE for $\boldsymbol{\beta}$ for small sample sizes.

It is worth noting that in some applications, it is hard to sample from the posterior density without resorting to SMC methods. This is because constructing a conditional distribution of the variable at the current time given previous ones is much easier than constructing the conditional distribution of that variable given all other variables. One example is the Lévy-

driven stochastic volatility model (Barndorff-Nielsen and Shephard, 2001), which is very popular in financial econometrics. In such cases, if one wishes to perform Bayesian analysis, then the PGS method is more desirable in these contexts.

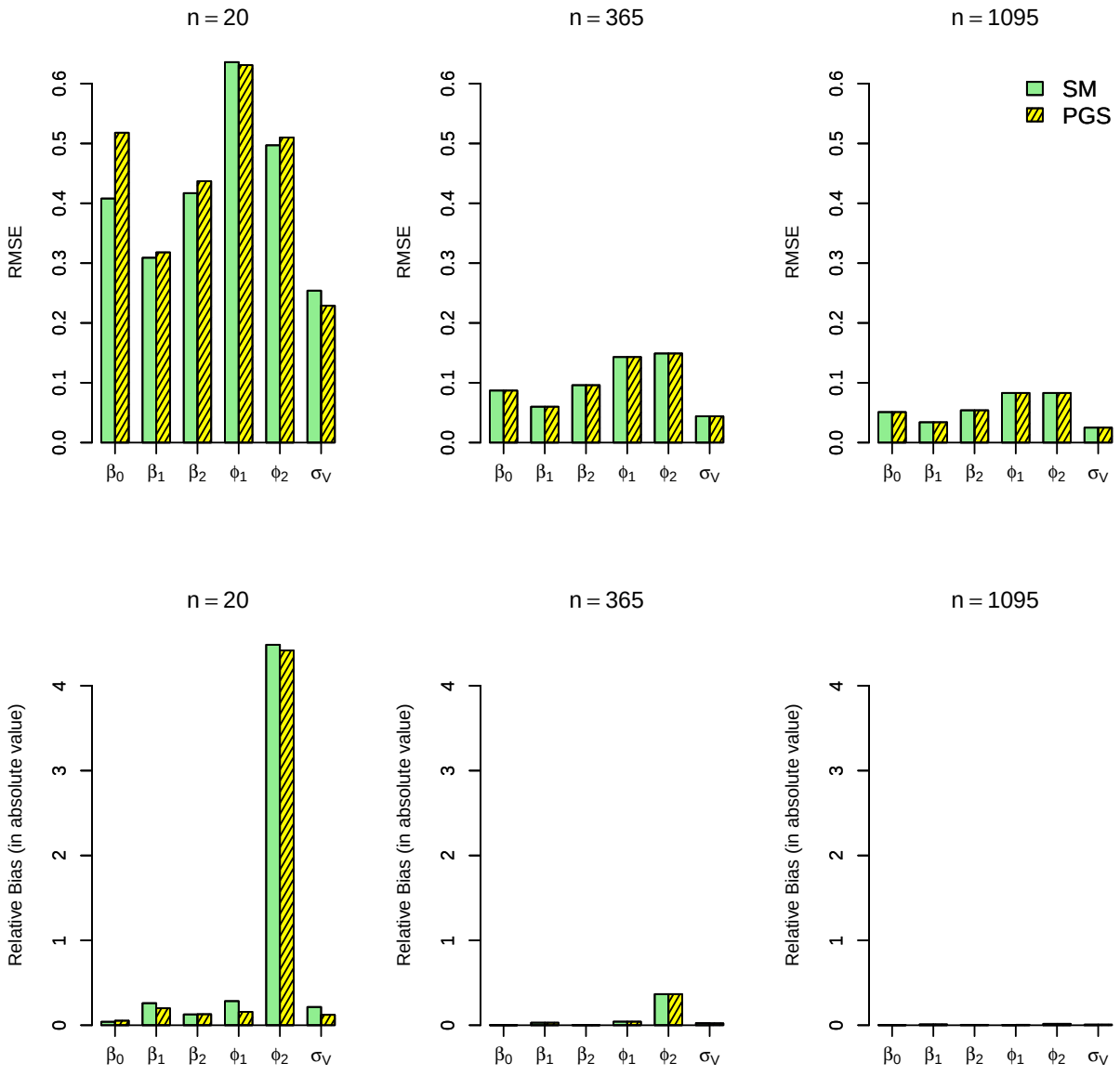

Figure 3.6: RMSE and relative bias (in absolute value) of parameter estimates for the SM and the PGS methods with $AR(2)$ latent process and weak serial correlation; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.06)$ and $\sigma_V = 0.4$.

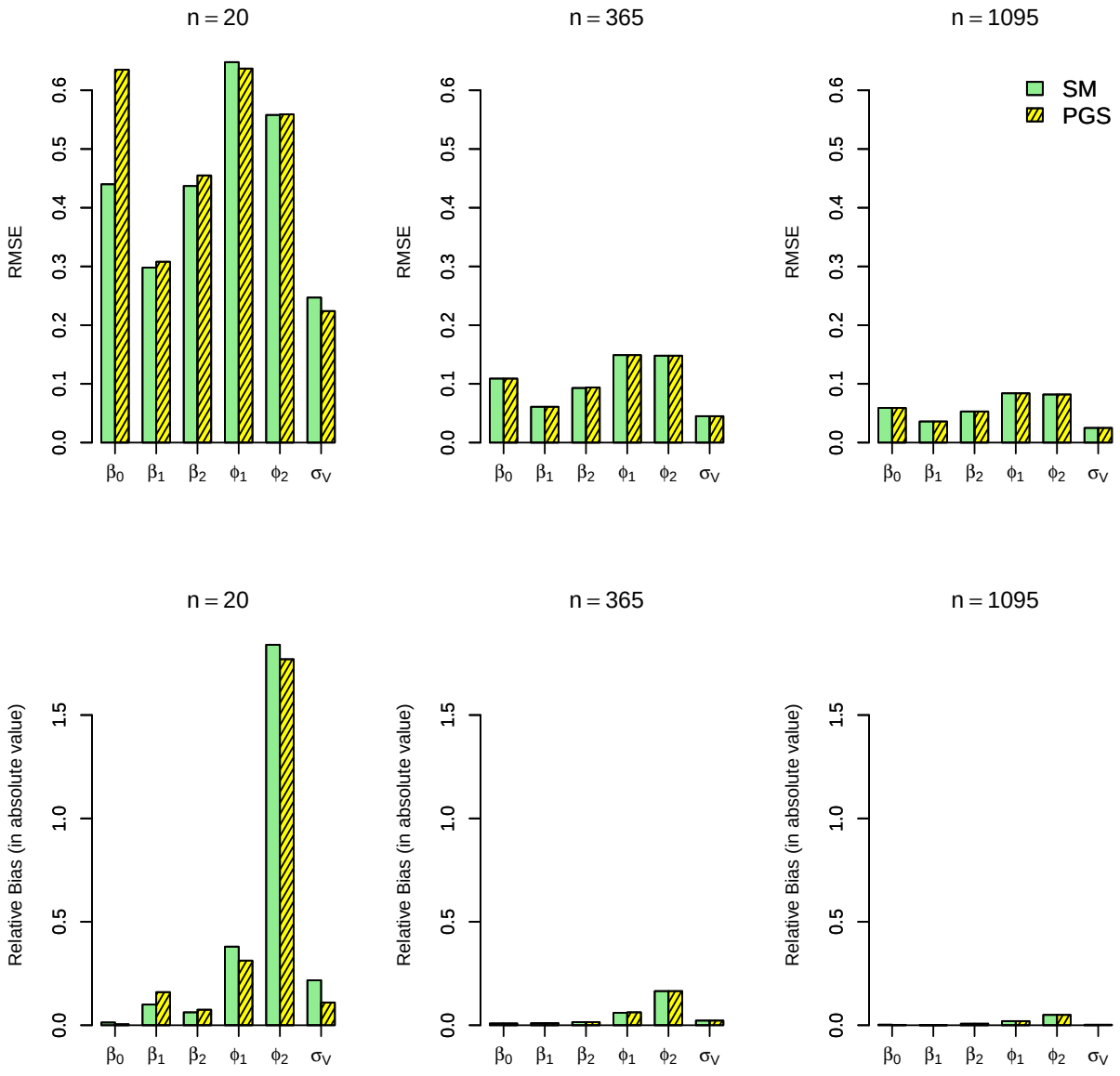

Figure 3.7: RMSE and relative bias (in absolute value) of parameter estimates for SM and PGS methods with $AR(2)$ latent process and moderate serial correlation; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $\sigma_V = 0.4$.

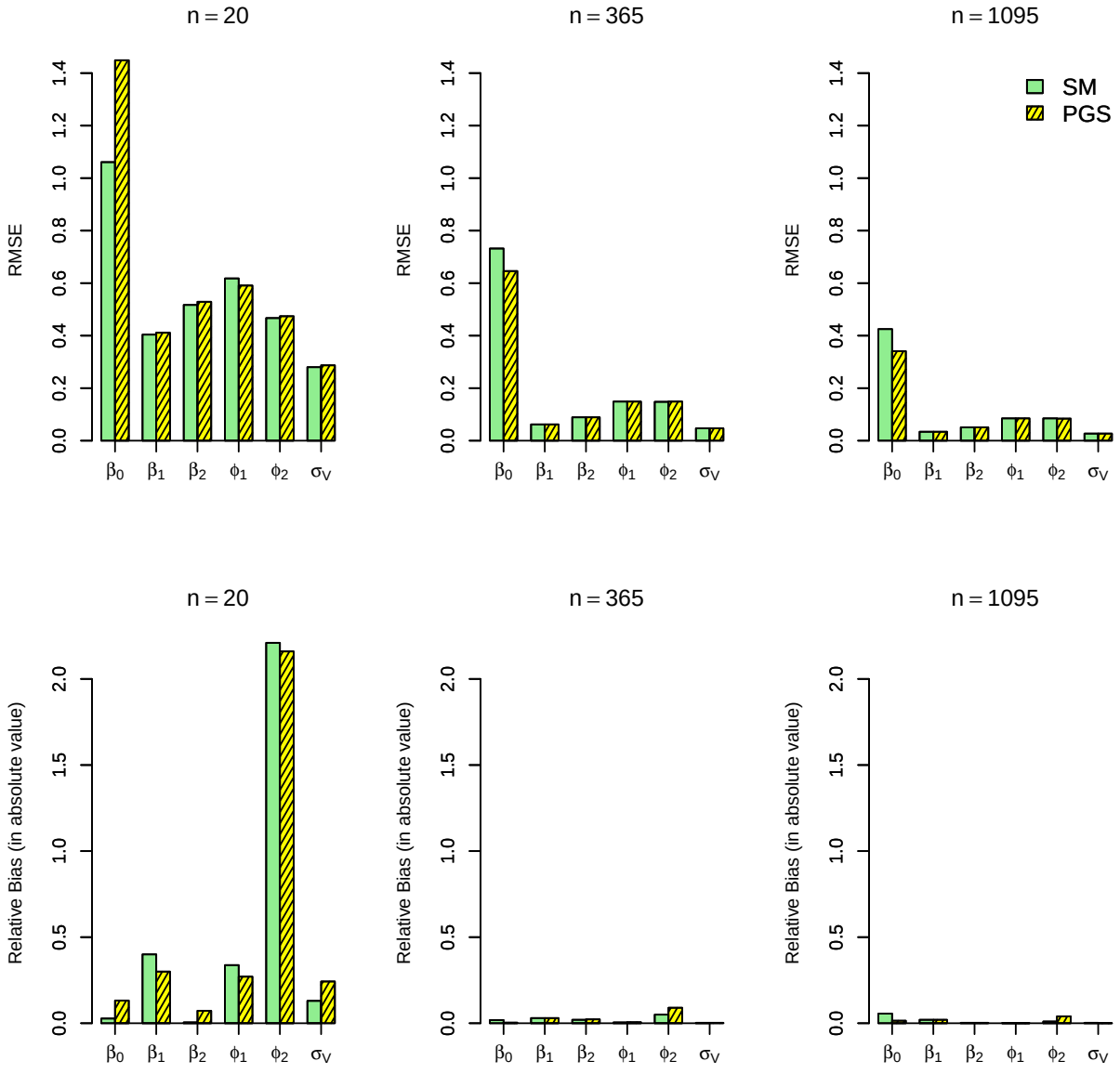

Figure 3.8: RMSE and relative bias (in absolute value) of parameter estimates for the SM and the PGS methods with $AR(2)$ latent process and strong serial correlation; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $\sigma_V = 0.4$.

Chapter 4

Bayesian Estimation Using Composite Likelihood

4.1 Motivation

The motivation for considering Bayesian composite likelihood methods is that, in many modern applications, the use of the full likelihood is problematic. This is due to the fact that its formula is unknown or the fact that it is not available in closed form. One example of the first difficulty is the max-stable process (Smith, 1990), processes that are used to construct probability models for rare events, for which the full likelihood is not available and only bivariate marginal densities available. The second difficulty arises in SSM or high dimensional missing data problems, for which the likelihood involves high dimensional integrals. In such situations, if one wishes to perform Bayesian analysis, the composite likelihood offers a possible surrogate to the full likelihood. The posterior distribution that results from replacing the full likelihood by the composite likelihood will be called the composite posterior distribution. Bayesian inference using the composite posterior is still at an exploratory stage and there has been a very limited number of papers on this topic. Most work was motivated by the desire to avoid using or specifying the full likelihood. For example, Smith and Stephenson (2009) replaces the full likelihood by a pairwise likelihood in MCMC methods to fit a max-stable process. Friel (2012) compared different versions of the composite likelihood to perform Bayesian analysis for a Gibbs random field. Related work to Bayesian inference with composite likelihood can be found in Monahan and Boos (1992) and Lazar (2003), where they used a likelihood that is not the conditional density of the data given the parameters in place

of the full likelihood to perform Bayesian analysis.

Smith and Stephenson (2009) noticed that the use of pairwise likelihood in place of the full likelihood in Bayesian inference may lead to an imprecise inference. That is, the variance in the composite posterior distribution is smaller than the truth and thus this will lead to a falsely narrower credible intervals. However, they did not mention how to resolve this problem. One possible explanation for this behavior of the composite posterior is that the composite likelihood treats the pairs of observations as being mutually independent and uses each observation many times. This can be seen as replacing the full likelihood by a version where the likelihood is raised to some power; which in turn will lead to a more concentrated posterior or, equivalently, less variability in the posterior samples. Three different adjustment methods have been proposed to address this issue. These adjustment methods aim to produce credible intervals constructed from the empirical quantiles of the adjusted draws that have approximately the nominal coverage. The first method was independently proposed by both Pauli et al. (2011) and Ribatet et al. (2012), and was termed magnitude adjustment by the latter. The rationale behind this method is to rescale the composite likelihood by raising it to a power, less than one, to achieve the proper magnitude. The second method was proposed by Ribatet et al. (2012) and was termed curvature adjustment. This method seeks to achieve the aim by stretching the horizontal axis of the log composite likelihood so as to match the asymptotic log density of the maximum composite likelihood estimator. Both of these methods share the fact that they are applied within MCMC iterations. By contrast, Shaby (2014) proposed a third adjustment method which is applied outside the MCMC iterations. He called this method the open faced sandwich (OFS) adjustment. The OFS method was inspired by the fact that the asymptotic distribution of the composite Bayes estimate has different covariance matrix than the limiting composite posterior distribution. Therefore, this method suggests rescaling the composite posterior samples to achieve the desired covariance matrix.

The aim of this chapter is to exploit the use of the composite likelihood in Bayesian inference of Poisson parameter-driven models. In the case of Poisson parameter driven models,

computing the full likelihood is computationally prohibitive, as it involves a high dimensional integral over the joint distribution of the latent variables $\{\boldsymbol{\eta}\}$. In the previous chapters, we discussed using the MCMC and PMCMC sampling to sample the latent variables. However, sampling latent variables may not have additional benefits. In this Chapter, we will consider Bayesian analysis of Poisson parameter-driven models using the composite likelihood method. We will focus on one version of the composite likelihood that is called the *pairwise composite likelihood* of order m, the product of the joint distributions of pairs that are not too far apart.

The remainder of this chapter is organized as follows. In Section 4.2, we will introduce the composite likelihood function. Section 4.3 introduces the three adjustment methods, the magnitude method in Section 4.3.1, the curvature method in 4.3.2, and the OFS method in Section 4.3.3. We will illustrate how to implement these adjustments within the M-H and Gibbs sampler algorithms in Section 4.4. Applying these adjustment methods in Poisson parameter-driven models will be considered in Section 4.5. A simulation study is considered in Section 4.6 for comparing the performance of the three adjustment methods. The methods were compared with the SM method to see which one provide closer results to the SM method.

4.2 Composite Likelihood

Composite likelihoods are pseudo likelihoods formed by composing low dimensional likelihood objects. Many composite likelihoods have been proposed in the literature under different names, such as Besag (1974) pseudo likelihood, Cox (1975) partial likelihood, the mth-order likelihood (Azzalini, 1983) and split data likelihood (Ryden, 1994). All of these can be seen as a special case of the composite likelihoods defined by Lindsay (1988).

As defined by Jin (2010), the composite likelihood refers to a likelihood objects formed by multiplying individual component likelihoods that corresponds to marginal or conditional events, formulated, for a random variable Y with density function $f(y;\boldsymbol{\theta})$, as

$$CL\left(\boldsymbol{\theta};\mathbf{y}\right)=\prod_{k\in K}L_k(\boldsymbol{\theta};y)^{w_k},$$

where $L_k(\boldsymbol{\theta}; y) = f(\{y \in E_k\}; \boldsymbol{\theta})$ is the likelihood function for an event E_k, where $\{E_k, k \in K\}$ is a set of events, and w_k are non-negative weights to be chosen. If the weights are all equal, then they can be ignored, and this is the composite likelihood under the independence of E_k's. See Varin et al. (2011) for a recent review of the topic.

Despite the fact that composite likelihoods are under-specified likelihoods, they have been receiving increasing attention. In some applications, the computation of the full likelihood is infeasible, due primarily to two reasons. The first reason is when the full likelihood is known but is too difficult to compute. The second reason is when the full likelihood is analytically unavailable. The first case occurs when the full likelihood involves high-dimensional integral, such as in state space models, and the second one arises in max-stable processes. These processes are used to build probability models for complex rare events for which closed forms are usually available only for the bivariate marginal densities. To circumvent these and other similar difficulties, composite likelihood approaches could be used; such approaches will leverage the high dimensional integral with low dimensional ones, aiming to reduce the computational difficulty so that it is possible to deal with large data sets and very complex models. Thus, the main reasons for using composite likelihood methods are that the computational complexity will be substantially reduced, and inferences based on composite likelihood methods have theoretical properties similar to those based on the full likelihood methods. Moreover, the relative efficiency of the maximum composite likelihood estimate relative to the maximum likelihood estimate is found to be high in many applications. Although a loss of efficiency is possible in some cases, in exceptional cases – see for example Mardia et al. (2007) and Jin (2010) – the maximum composite likelihood estimate is fully efficient.

The maximum composite likelihood estimator (MCLE), denoted by $\hat{\boldsymbol{\theta}}_c$, locates the maximum of composite log-likelihood $c\ell(\boldsymbol{\theta}; \mathbf{y}) = \log CL(\boldsymbol{\theta}; \mathbf{y})$ and can be found by solving the composite score function $u_c(\boldsymbol{\theta}; \mathbf{y}) = \nabla c\ell(\boldsymbol{\theta}; \mathbf{y})$. It was shown that the MCLE is asymptotically normally distributed (see Lindsay (1988) and Kent (1982)), that is

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) \xrightarrow{d} N_d\left(0, G^{-1}(\boldsymbol{\theta})\right),$$

where $N_d(\mu, \Sigma)$ is the $d-$dimensional normal distribution with mean μ and variance Σ. The

matrix $G(\boldsymbol{\theta})$ is known as the Godambe information matrix in a single observation and is given by:

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}H(\boldsymbol{\theta}), \tag{4.1}$$

where the sensitivity matrix $H(\boldsymbol{\theta})$ is given by

$$H(\boldsymbol{\theta}) = E\{-\nabla u_c(\boldsymbol{\theta}, y)\},$$

and the variability matrix $J(\boldsymbol{\theta})$ is given by

$$J(\boldsymbol{\theta}) = Var\{u_c(\boldsymbol{\theta}, y)\}.$$

Both the expectation and the variance above are with respect to the true density.

Two well known examples of composite likelihood that are relevant to the version used in this thesis are the pairwise composite likelihood (Cox and Reid, 2004) and the pairwise composite likelihood of order m. Each is defined in the following.

The pairwise composite likelihood is constructed by all possible pairs of observations, and is given by

$$CL_p(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} f_{i,j}(y_i, y_j; \boldsymbol{\theta}).$$

There are a number of situations where the dependence is believed to be negligible between pairs of observations that are more than m units far apart. In these cases, incorporating these pairs in the composite likelihood will increase the computational cost and may reduce the efficiency of the estimator. For an example, see Varin and Vidoni (2009) and Apanasovich et al. (2008). In these cases, it is encouraged to include pairs that are at most m units far apart. The resulting composite likelihood is called the pairwise composite likelihood of order m, and is given by

$$CL_p^{(m)}(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^{n-m} \prod_{j=1}^{m} f_{i,i+j}(y_i, y_{i+j}; \boldsymbol{\theta}).$$

In the case of the composite likelihood of order m, there is no general procedure how to determine m. However, for the case of Poisson model, Ng et al. (2011) found through simulation studies that $m = 3$, or 4 is good enough to have statistically efficient estimates relative to the maximum likelihood estimates (MLE) in the frequentist paradigm.

4.3 Adjustment of The Composite Likelihood

4.3.1 Magnitude Adjustment

Magnitude adjustment is independently proposed by Pauli et al. (2011) and Ribatet et al. (2012) and is inspired by the following fact. The likelihood ratio statistic, $\Lambda(\boldsymbol{\theta}_0)$, for testing the hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ has a limiting distribution that is the chi-Square distribution with d degrees of freedom, denoted by χ^2_d, where d is the dimension of the parameter $\boldsymbol{\theta}$. In contrast, the composite likelihood ratio statistic, $\Lambda_c(\boldsymbol{\theta}_0)$, has a non-standard asymptotic distribution, which is the distribution of $\sum_{i=1}^{d} \lambda_i X_i$ (Kent, 1982), where $\lambda_i, i = 1, ..., d$ are the eigenvalues of the matrix $H(\boldsymbol{\theta}_0)^{-1} J(\boldsymbol{\theta}_0)$ and $X_1, ..., X_d$ are independent χ^2_1 random variables. Different adjustments have been proposed to the composite likelihood so as to have $\Lambda_c(\boldsymbol{\theta}_0)$ approximately adheres to a standard asymptotic distribution χ^2_d.

Geys et al. (1999) proposed an adjustment to the composite likelihood by selecting a suitable weight to its components. The selected weight was $1/\bar{\lambda}$, where $\bar{\lambda}$ is the average of the eigenvalues λ_i. When $\boldsymbol{\theta}$ is scalar, i.e., $d = 1$, this adjustment causes the composite likelihood ratio statistic to follow a χ^2_1 as an asymptotic distribution. For $d > 1$, $\Lambda_c(\boldsymbol{\theta}_0)$ approximately follows a χ^2_d as an asymptotic distribution. In fact, only the mean of the $\Lambda_c(\boldsymbol{\theta}_0)$ coincides with that of the χ^2_d, while other moments may differ.

Pauli et al. (2011) and Ribatet et al. (2012) proposed to use the Geys et al. (1999) adjustment to the composite likelihood when being used in Bayes' formula in place of the full likelihood. They showed that the adjusted composite posterior converges to the asymptotic distribution of the maximum composite likelihood estimator $\hat{\boldsymbol{\theta}}_c$ when $\boldsymbol{\theta}$ is scalar. In the case of vector $\boldsymbol{\theta}$, the adjusted composite posterior offers improvement over the unadjusted one in the sense that the variability in the adjusted posterior is better approximating the asymptotic variance of $\hat{\boldsymbol{\theta}}_c$. The adjusted composite posterior using the magnitude adjustment is given by

$$\pi_{mag}(\boldsymbol{\theta}|y) \propto \{CL(\boldsymbol{\theta};\mathbf{y})\}^{1/\bar{\lambda}}\pi(\boldsymbol{\theta}). \tag{4.2}$$

4.3.2 Curvature Adjustment

The curvature adjustment method is inspired by the same fact that $\Lambda(\boldsymbol{\theta}_0)$ has a χ^2_d as an asymptotic distribution while $\Lambda_c(\boldsymbol{\theta}_0)$ does not; however, it uses a different strategy to ensure this convergence. The curvature adjustment method stretches the curvature of the composite log likelihood so as to asymptotically match the log density of $\hat{\boldsymbol{\theta}}_c$, which in turn leads to a standard χ^2_d asymptotic distribution of $\Lambda_c(\boldsymbol{\theta}_0)$. The rationale behind that is the following; the likelihood ratio statistic, $\Lambda(\boldsymbol{\theta}_0)$, has an asymptotic distribution of χ^2_d because the second derivative of the log likelihood function evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, $-\frac{1}{n}\nabla^2\ell(\hat{\boldsymbol{\theta}};\mathbf{y})$, converges almost surely to the inverse of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$. Therefore, in curvature adjustment, Ribatet et al. (2012) wanted to ensure that the second derivative of the composite log likelihood function evaluated at the maximum composite likelihood estimator $\hat{\boldsymbol{\theta}}_c$, $-\frac{1}{n}\nabla^2\ell_c(\hat{\boldsymbol{\theta}}_c;\mathbf{y})$, converges almost surely to the inverse of the asymptotic covariance matrix of the maximum composite likelihood estimator, i.e., $G^{-1} = H^{-1}(\boldsymbol{\theta})J(\boldsymbol{\theta})H^{-1}(\boldsymbol{\theta})$. This was achieved by replacing $\boldsymbol{\theta}$ in the composite likelihood by $\boldsymbol{\theta}^{adj} = \hat{\boldsymbol{\theta}}_c + C(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_c)$, where C is a semi-negative definite matrix such that

$$C^T H(\boldsymbol{\theta}_0)C = H(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}H(\boldsymbol{\theta}_0). \tag{4.3}$$

One choice of C is

$$C = M^{-1}M_A,$$

where M and M_A are the square roots of the matrices $H(\boldsymbol{\theta}_0)$ and $H(\boldsymbol{\theta}_0)J(\boldsymbol{\theta}_0)^{-1}H(\boldsymbol{\theta}_0)$, respectively. The matrix square roots are not unique, and possible methods are Cholesky square roots or singular value decomposition. The latter method is recommended by Ribatet et al. (2012) to ensure that any directions of symmetry is preserved by the mapping in Eq. (4.3). This adjustment to the composite likelihood and the choice of the matrix C

above were proposed by Chandler and Bate (2007) in the context of cluster data analysis using independence likelihood, which is the likelihood resulting from assuming independent observations.

Ribatet et al. (2012) showed that the resulting composite posterior converges to the asymptotic distribution of the maximum composite likelihood estimator. The adjusted composite posterior using curvature adjustment will be defined as follows

$$\pi_{curv}(\boldsymbol{\theta}|\mathbf{y}) \propto CL(\boldsymbol{\theta}^{adj};\mathbf{y})\pi(\boldsymbol{\theta}). \tag{4.4}$$

4.3.3 Open Faced Sandwich Adjustment

The inspiring fact that motivates the OFS is that the composite Bayes estimator $\hat{\boldsymbol{\theta}}_{CB}$, typically the mean of the composite posterior distribution, is asymptotically normally distributed with a mean equal to the true parameter $\boldsymbol{\theta}_0$ and variance $G^{-1}(\boldsymbol{\theta}_0)$, (see Chernozhukov and Hong (2003) for details). This result is also true for the maximum composite likelihood estimator. However, Chernozhukov and Hong (2003) observed that the limiting composite posterior distribution is normal with $H^{-1}(\boldsymbol{\theta}_0)$ as a covariance matrix, which is different from the covariance matrix of the asymptotic distribution of the composite Bayes estimator, G^{-1}. Shaby (2014) suggests sampling from the unadjusted composite posterior then pre-multiplying the samples, after centering, by an estimator $\hat{\Omega}$ of the matrix $\Omega = H^{-1}(\boldsymbol{\theta}_0)J^{\frac{1}{2}}(\boldsymbol{\theta}_0)H^{\frac{1}{2}}(\boldsymbol{\theta}_0)$, as in Eq. (4.5) below, to achieve the desired covariance matrix

$$\boldsymbol{\theta}^{(j)}_{OFS} = \hat{\boldsymbol{\theta}}_{CB} + \hat{\Omega}(\boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}}_{CB}), \tag{4.5}$$

where $\boldsymbol{\theta}^{(j)}$ and $\boldsymbol{\theta}^{(j)}_{OFS}$ are the unadjusted and the adjusted samples, respectively. It is easy to show that the covariance matrix of the adjusted samples is now G^{-1} as desired. To estimate the matrix Ω, we need to estimate both the matrix H and J, which will be discussed in the next section.

4.3.4 Estimating H and J

In order to perform the adjustment methods mentioned earlier, we need to estimate the matrices H and J. Estimation of the sensitivity matrix H is simpler due to the fact that the second Bartlett identity (Ferguson, 1996), $H(\boldsymbol{\theta}) \neq J(\boldsymbol{\theta})$, remains valid for each likelihood component forming the composite likelihood. Thus, a sample estimate of the matrix H is given by

$$\hat{H}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \nabla u_c \left(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}_k \right),$$

where $u_c \left(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}_k \right)$ is the score function that corresponds to the k^{th} likelihood component forming the composite likelihood.

Greater difficulties arise in estimating the matrix J, as the second Bartlett identity does not hold for the composite likelihood. This is due to independent assumption among the likelihood terms forming the composite likelihood, which does not hold in practice. When a single time series is observed, estimation of the matrix J is possible when independent or pseudo-independent replicates of the data are available (Varin et al., 2011). In cases where the data can be grouped into B pseudo-independent subgroups, $G_1, \ldots, G_B$, a sample estimate of the matrix J is given by

$$\hat{J}(\boldsymbol{\theta}) = \frac{1}{B} \sum_{b=1}^{B} n_b u_c \left(\hat{\boldsymbol{\theta}}_{CB}; y \in G_b \right) u_c \left(\hat{\boldsymbol{\theta}}_{CB}; y \in G_b \right)^T,$$

where n_b is the size of group b. This method was used by Heagerty and Lele (1998) in the context of binary spatial data and termed by them window subsampling. In situations where window subsampling is not valid but it is possible to simulate the process that generated $\mathbf{y}$, estimation of the matrix J can be accomplished using Monte Carlo integration – also called parametric bootstrapping – as follows. Let $\mathbf{y}_1, \ldots, \mathbf{y}_S$ be S independent replicates, generated under $\hat{\boldsymbol{\theta}}_{CB}$, of the process that generated $\mathbf{y}$. Then an estimator of the matrix J is given by

$$\hat{J}(\boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^{S} u_c(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}^{(s)}) u_c(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}^{(s)})^T. \tag{4.6}$$

It is also possible to estimate the matrix H using the same method as follows

$$\hat{H}(\boldsymbol{\theta}) = \frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K} u_c(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}^{(s)} \in E_k) u(\hat{\boldsymbol{\theta}}_{CB}; \mathbf{y}^{(s)} \in E_k)^T. \tag{4.7}$$

In this thesis, we will use the Monte Carlo integration method to estimate both matrices H and J, as in Eq. (4.7) and Eq. (4.6).

4.4 MCMC Using Composite Likelihood

Usually it is hard to sample from the composite posterior, adjusted or not, without resorting to MCMC methods, such as the Gibbs sampler or the M-H algorithm. The reason is that the composite posterior involves integral that can not be solved in closed form; thus it making it difficult to sample from the composite posterior using an inverse distribution function or accept-reject methods. In this section, we will give a general discussion on how to implement MCMC methods using composite likelihood to perform Bayesian inference. A more specific discussions for a Poisson parameter driven model will be given in Section 4.5. In the following two sections, we summarize the adjusted M-H and Gibbs sampler algorithms as described by Ribatet et al. (2012) and Shaby (2014).

4.4.1 Adjusted Metropolis Hastings Algorithm

Let $\pi_c(\boldsymbol{\theta}|\mathbf{y})$ be the unadjusted composite posterior. Then the M-H algorithm using $\pi_c(\boldsymbol{\theta}|\mathbf{y})$ can be implemented as in Algorithm I below.

Algorithm I:

- Input: Unadjusted composite posterior $\pi_c(\boldsymbol{\theta}|\mathbf{y})$, a proposal density $q(.|\boldsymbol{\theta})$ and initial value $\boldsymbol{\theta}^{(0)}$.
- Output: A realization of $\boldsymbol{\theta}^{(t)}$, $t = 1, \ldots, T$.
- For $t = 1, ..., T$ do the following:
    1. Sample a value $\boldsymbol{\theta}^{(*)}$ from $q(.|\boldsymbol{\theta}^{(t-1)})$.

2. Calculate

$$\alpha(\boldsymbol{\theta}^{(*)}, \boldsymbol{\theta}^{(t-1)}) = min\left\{1, \frac{\pi_c(\boldsymbol{\theta}^{(*)}|\mathbf{y}) \times q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})}{\pi_c(\boldsymbol{\theta}^{(t-1)}|\mathbf{y}) \times q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}\right\}.$$

3. Sample $U \sim \text{Uniform}(0, 1)$.

4. Let

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^{(*)}, & \text{If } U \leq \alpha\left(\boldsymbol{\theta}^{(*)}, \boldsymbol{\theta}^{(t-1)}\right) \\ \boldsymbol{\theta}^{(t-1)}, & \text{otherwise} \end{cases}.$$

For the OFS adjustment we apply Eq. (4.5) to $\boldsymbol{\theta}^{(0)}, ..., \boldsymbol{\theta}^{(T)}$ to get the adjusted samples. The Bayesian inference can be carried out using the adjusted samples $\{\boldsymbol{\theta}^{(0)}_{OFS}, ..., \boldsymbol{\theta}^{(T)}_{OFS}\}$.

For the magnitude or the curvature adjustments, we need to run Algorithm I twice. In the first time, we run it as it is using $\pi_c(\boldsymbol{\theta}|\mathbf{y})$ to get $\hat{\boldsymbol{\theta}}_{CB}$, and then use it to estimate the matrices $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ and then calculating the necessary adjustments. In the second time, we run it using $\pi_{mag}(\boldsymbol{\theta}|\mathbf{y})$ or $\pi_{curv}(\boldsymbol{\theta}|\mathbf{y})$ in place of $\pi_c(\boldsymbol{\theta}|\mathbf{y})$ to take into account the magnitude or the curvature adjustment, respectively.

4.4.2 Adjusted Gibbs Sampler

In situations where the dimension of $\boldsymbol{\theta}$ is high, it may be hard to choose a high dimensional proposal to apply the M-H algorithm, or the probability of acceptance for the M-H algorithm may be too low. In such situations, the parameters of interest are often partitioned into low dimensional blocks, and the Gibbs sampler is employed by successively sampling from the full conditional posteriors of each block given all other blocks, as follows.

Let us partition $\boldsymbol{\theta}$ into B blocks, $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_B$. If we wish to draw samples from the joint composite posterior

$$\pi_c(\boldsymbol{\theta}|y) \propto CL(\boldsymbol{\theta}; \mathbf{y})\pi(\boldsymbol{\theta}),$$

then the Gibbs sampler is implemented by successively drawing from

$$\pi_c(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b}, \mathbf{y}) \propto CL(\boldsymbol{\theta}_b; \mathbf{y})\pi(\boldsymbol{\theta}_b), \quad b = 1, ..., B,$$

where $\boldsymbol{\theta}_{-b}$ are all other blocks, but not the current sampled block $\boldsymbol{\theta}_b$. It is usually hard to separate $\boldsymbol{\theta}_b$ from $\boldsymbol{\theta}$ in the composite likelihood, so $CL(\boldsymbol{\theta}_b; \mathbf{y})$ and $CL(\boldsymbol{\theta}, \mathbf{y})$ have usually the same expression, but $CL(\boldsymbol{\theta}_b; \mathbf{y})$ is seen as a function of $\boldsymbol{\theta}_b$ where all other blocks are held fixed.

Ribatet et al. (2012) proposed two versions of the Gibbs sampler when the composite posterior is used. The first one is the overall Gibbs sampler, and the second is the adaptive Gibbs sampler. The key difference between the two versions is that in the overall Gibbs sampler, the matrices H and J – and thus the adjustments – are estimated only one time at the beginning of the Gibbs sampler, and then it proceeds as usual. By contrast, in the adaptive Gibbs sampler, the aforementioned matrices are estimated in each iteration of the Gibbs sampler for each block of the parameters given other blocks held fixed at the newly sampled values.

The adaptive Gibbs sampler is more computationally expensive – potentially a thousand fold – compared to the overall Gibbs sampler. However, Ribatet et al. (2012) observed via simulation studies that there are very minor differences in the results between the two versions of the Gibbs sampler when the curvature adjustment is used, while the adaptive Gibbs sampler secures improvement over the overall one when using magnitude adjustment.

In the case of the OFS adjustment method, Shaby (2014) distinguished between two situations when applying the Gibbs sampler. The first situation is when the posteriors for all blocks, $\{\boldsymbol{\theta}_b, b = 1, ..., B\}$, is a composite posterior; in this case the Gibbs sampler can be run by successively drawing from $\pi_c(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b}; \mathbf{y})$, $b =, 1..., B$, and then apply OFS adjustment post hoc, as in the M-H algorithm. The second situation arises when there are some blocks $\boldsymbol{\theta}_b$, for some $b = 1, .., B$, for which the posterior is free of the composite likelihood and only is proportional to the prior, i.e, $\pi_c(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b}, \mathbf{y}) \propto \pi(\boldsymbol{\theta}_b)$. In this case, the OFS adjustment must occur within the sampling algorithm as follows.

Suppose the posterior for the block $\boldsymbol{\theta}_b$ is a composite posterior, while the posterior for the block $\boldsymbol{\theta}_{b+1}$ is not. At iteration i of the Gibbs sampler, sample $\boldsymbol{\theta}_b^{(i)}$ from $\pi_c(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b}, \mathbf{y})$, then apply the OFS to the sample to get $\boldsymbol{\theta}_{b(OFS)}^{(i)}$, then use this to sample $\boldsymbol{\theta}_{b+1}^{(i)}$ from $\pi_c(\boldsymbol{\theta}_{b+1}|\boldsymbol{\theta}_{-(b+1)}, \mathbf{y})$.

Now, $\boldsymbol{\theta}_{b+1}^{(i)}$ will be used without any adjustment to sample the block $\boldsymbol{\theta}_{b+2}^{(i)}$.

4.5 Poisson Parameter Driven Model

4.5.1 The Metropolis-Hastings Algorithm

As mentioned earlier, we will study the use of pairwise composite likelihood of order m in place of the full likelihood in Bayes' formula to get the composite posterior and perform Bayesian analysis based on this composite posterior density. A composite likelihood of order m for a Poisson parameter-driven model is defined as follows

$$CL(\boldsymbol{\theta};\mathbf{y}) = \prod_{i=1}^{n-m}\prod_{j=1}^{m}\int\int g(y_i|\eta_i,\boldsymbol{\theta})g(y_{i+j}|\eta_{i+j},\boldsymbol{\theta})f_j(\eta_i,\eta_{i+j}|\boldsymbol{\theta})d\eta_i d\eta_{i+j},$$

where

$$g(y_l|\eta_l,\boldsymbol{\theta}) = \exp\left\{-e^{\boldsymbol{\beta}^T\mathbf{X}_l+\eta_l} + y_l\left(\boldsymbol{\beta}^T\mathbf{X}_l+\eta_l\right) - \log\Gamma(y_l+1)\right\},$$

and $f_j(\eta_i,\eta_{i+j};\boldsymbol{\theta})$ is the bivariate normal density with mean $\mathbf{0}$ and covariance matrix

$$\Sigma_j = \begin{bmatrix} \gamma_0 & \gamma_j \\ \gamma_j & \gamma_0 \end{bmatrix}.$$

Thus, the composite posterior density of $\boldsymbol{\theta}= \{\boldsymbol{\beta}, r_1^*,\ldots,r_p^*, W\}$, where $W=\log(\sigma_V)$, is given by

$$\pi_c\left(\boldsymbol{\theta}|\mathbf{y}\right) \propto CL\left(\boldsymbol{\theta};\mathbf{y}\right) \times \prod_{k=0}^{K}\varphi_N(\beta_k;\mu_\beta,\sigma_\beta^2) \times \prod_{k=1}^{p}\pi\left(r_k^*\right) \times \pi\left(w\right).$$

Sampling $\boldsymbol{\theta}$ from the above unadjusted composite posterior $\pi_c\left(\boldsymbol{\theta}|\mathbf{y}\right)$ can be accomplished using the M-H algorithm with independent normal proposal densities as follows:

1. Sample $\boldsymbol{\theta}^{(*)}$ from $N(\boldsymbol{\theta}^{(t-1)},D)$, where $\boldsymbol{\theta}^{(t-1)}$ is the previously accepted value of $\boldsymbol{\theta}$ and D is a diagonal matrix.

2. Sample u from U(0,1).

3. Calculate $\alpha=\dfrac{\pi_c\left(\boldsymbol{\theta}^{(*)} \mid \mathbf{y}\right)}{\pi_c\left(\boldsymbol{\theta}^{(t-1)} \mid \mathbf{y}\right)}$.

4. Let

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^{(*)}, & \text{If } U \leq \alpha \\ \boldsymbol{\theta}^{(t-1)}, & \text{otherwise} \end{cases}.$$

In the case of the OFS adjustment method, we will perform Bayesian analysis based on the adjusted posterior samples obtained by applying Eq. (4.5) to the unadjusted posterior outputs. In the case of the magnitude and curvature adjustment methods, the above algorithm will be run twice. First, it will be run using the unadjusted composite likelihood to find $\hat{\boldsymbol{\theta}}_c$ and then estimate the matrices H and J. Second, it will be run again and the adjusted posteriors, $\pi_{mag}(\boldsymbol{\theta}|\mathbf{y})$ and $\pi_{curv}(\boldsymbol{\theta}|\mathbf{y})$, will be used in place of $\pi_c(\boldsymbol{\theta}|\mathbf{y})$ to take into account the magnitude and curvature adjustment methods, respectively.

4.5.2 Gradients of the Log Composite Likelihood

For the evaluation of the matrices $\hat{H}$ and $\hat{J}$, we need the gradient of the log composite likelihood $c\ell(\boldsymbol{\theta};\mathbf{y})$ with respect to $\boldsymbol{\theta}$. These derivatives can be found in Ng et al. (2011), which we will now summarize.

The log composite likelihood $c\ell(\boldsymbol{\theta};\mathbf{y})$ can be written as

$$\begin{aligned} c\ell(\boldsymbol{\theta};\mathbf{y}) &\propto \sum_{i=1}^{n-m}\sum_{j=1}^{m} \log\left(\int\int h(y_i, y_{i+j}, \eta_i, \eta_{i+j}) f_j(\eta_i, \eta_{i+j};\boldsymbol{\theta}) d\eta_i d\eta_{i+j}\right) \\ &= \sum_{i=1}^{n-m}\sum_{j=1}^{m} \log\left(E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}\right), \end{aligned}$$

where $E\{h\}$ stands for the expectation of h, and

$$h(y_i, y_{i+j}, \eta_i, \eta_{i+j}) = g(y_i|\eta_i, \boldsymbol{\theta}) g(y_{i+j}|\eta_{i+j}, \boldsymbol{\theta}).$$

The derivative of the composite log-likelihood is given by,

$$\frac{\partial c\ell(\boldsymbol{\theta};\mathbf{y})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n-m}\sum_{j=1}^{m} \frac{\partial \log\left(E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}\right)}{\partial \boldsymbol{\theta}},$$

using the chain rule, we have

$$\frac{\partial \log\left(E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}\right)}{\partial \boldsymbol{\theta}} = \frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \boldsymbol{\theta}} \times \frac{1}{E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}.$$

Derivative with respect to $\boldsymbol{\beta}$

$$\frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \beta_k} = \left(X_{ik}, X_{(i+j)k}\right) \Sigma_j^{-1} \begin{pmatrix} E\{\eta_i h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\} \\ E\{\eta_{i+j} h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\} \end{pmatrix}$$

Derivative with respect to γ_k

$$\frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \gamma_k} = \frac{1}{2} E\left\{ \operatorname{tr}\left\{ \frac{\partial \Sigma_j}{\partial \gamma_k} \left(\Sigma_j^{-1} \boldsymbol{\eta}\boldsymbol{\eta}^T \Sigma_j^{-1} - \Sigma_j^{-1} \right) \right\} h(y_i, y_{i+j}, \eta_i, \eta_{i+j}) \right\},$$

where $\operatorname{tr}\{\Sigma\}$ is the trace of the matrix Σ.

Derivative with respect to $(\sigma_V, \phi_1, \ldots, \phi_p)$

$$\frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \phi_k} = \frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \gamma_0} \frac{\partial \gamma_0}{\partial \phi_k} + \frac{\partial E\{h(y_i, y_{i+j}, \eta_i, \eta_{i+j})\}}{\partial \gamma_i} \frac{\partial \gamma_j}{\partial \phi_k}.$$

$$\frac{\partial E h(y_i, y_{i+j}, \eta_i, \eta_{i+j})}{\partial \sigma_V} = \frac{\partial E h(y_i, y_{i+j}, \eta_i, \eta_{i+j})}{\partial \gamma_0} \frac{\partial \gamma_0}{\partial \sigma_V} + \frac{\partial E h(y_i, y_{i+j}, \eta_i, \eta_{i+j})}{\partial \gamma_i} \frac{\partial \gamma_j}{\partial \sigma_V}.$$

4.5.3 Evaluating the Composite Likelihood and the Gradients

The evaluation of the composite likelihood and gradients requires the evaluation of the double integrals. This can be done by Laplace approximation. Laplace approximation is the fastest numerical integration method and it performs well in discrete mixed models, especially Poisson mixed models (Joe, 2008). The Laplace approximation involves evaluating the Hessian matrix. If the evaluation of Hessian matrix is not feasible, then other numerical integration methods need to be used to evaluate the double integrals in the composite likelihood, which will add more computational cost to the composite likelihood method.

To use Laplace approximation, we need to write the integrals in the composite likelihood function in the form

$$\mathrm{I}_1 = \int \int e^{-\xi(\eta_l, \eta_{l+k})} d\eta_l d\eta_{l+k}, \tag{4.8}$$

where $\xi(\eta_l, \eta_{l+k}) = -\log g\left(y_l|\eta_l, \boldsymbol{\theta}\right) - \log g\left(y_{l+k}|\eta_{l+k}, \boldsymbol{\theta}\right) - \log f_k(\eta_l, \eta_{l+k}|\boldsymbol{\theta})$.

Let $\hat{\boldsymbol{\eta}} = (\hat{\eta}_l, \hat{\eta}_{l+k})$ be a column vector that minimizes $\xi(\eta_l, \eta_{l+k})$. Then Laplace approximation of the integral above relies on Taylor series expansion of the function ξ about $\hat{\boldsymbol{\eta}}$. The Taylor expansion of ξ about $\hat{\boldsymbol{\eta}}$ up to the second order is given by

$$\xi(\boldsymbol{\eta}) \approx \xi(\hat{\boldsymbol{\eta}}) + \nabla\xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) + \frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})^T \nabla^2\xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}).$$

Using the fact that $\nabla\xi(\hat{\boldsymbol{\eta}}) = 0$, the integral I_1 in Eq. (4.8) can be approximated by

$$\begin{aligned} \mathrm{I}_1 &\approx e^{-\xi(\hat{\boldsymbol{\eta}})} \int\int e^{-\frac{1}{2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2\xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})} d\eta_l d\eta_{l+k} \\ &= 2\pi e^{-\xi(\hat{\boldsymbol{\eta}})} \left(\det\left[\nabla^2\xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}}, \end{aligned} \tag{4.9}$$

where $\det(\Sigma)$ is the determinant of the matrix Σ. The second line in Eq. (4.9) above stems from the fact that the expression being integrated in the first line of the equation is the density of a bivariate normal distribution without the normalizing constant, therefore the integral will be just the reciprocal of the normalizing constant.

The integrals in the gradients of the log composite likelihood with respect to the parameters can be written in the form

$$\mathrm{I}_2 = \int\int a(\boldsymbol{\eta}) \frac{e^{-\xi(\boldsymbol{\eta})}}{\int\int e^{-\xi(\boldsymbol{\eta})} d\eta_l d\eta_{l+k}} d\eta_l d\eta_{l+k}. \tag{4.10}$$

Here we have two cases with respect to the function $a(\boldsymbol{\eta})$. First, if $a(\boldsymbol{\eta})$ is a linear combination of η_l and η_{l+k}, then the Laplace approximation to the integral is $a(\hat{\boldsymbol{\eta}})$. This can be also seen by expanding $\xi(\boldsymbol{\eta})$ about $\hat{\boldsymbol{\eta}}$. Therefore, the integral I_2 in Eq. (4.10) can be approximated by

$$\begin{aligned} \mathrm{I}_2 &\approx \int\int a(\boldsymbol{\eta}) \frac{e^{-\frac{1}{2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2\xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})}}{2\pi \left(\det\left[\nabla^2\xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}}} d\eta_l d\eta_{l+k} \\ &= a(\hat{\boldsymbol{\eta}}). \end{aligned} \tag{4.11}$$

The first line of Eq. (4.11) can be seen as the expected value of the linear combination $a(\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ distributed as a bivariate normal with mean $\hat{\boldsymbol{\eta}}$; therefore, this integrates to $a(\hat{\boldsymbol{\eta}})$.

Second, if $a(\boldsymbol{\eta})$ is not a linear combination of $\boldsymbol{\eta}$, then the Laplace approximation to the integral I_2 in Eq. (4.10) is given by

$$a(\hat{\boldsymbol{\eta}}) + \frac{1}{2}\mathrm{tr}\left\{\left[\nabla^2\xi\left(\hat{\boldsymbol{\eta}}\right)\right]^{-1} \nabla^2 a\left(\hat{\boldsymbol{\eta}}\right)\right\}. \tag{4.12}$$

The formula in Eq. (4.12) above can be obtained by expanding both $a(\boldsymbol{\eta})$ and $\xi(\boldsymbol{\eta})$ about $\hat{\boldsymbol{\eta}}$. So, the integral I_2 in Eq. (4.10) can be approximated by

$$\begin{aligned}
\mathrm{I}_2 \approx \quad & \int\int a(\hat{\boldsymbol{\eta}}) \frac{e^{-\frac{1}{2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2 \xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})}}{2\pi \left(\det\left[\nabla^2 \xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}}} d\eta_l d\eta_{l+k} \\
& + \int\int \nabla a(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) \frac{e^{-\frac{1}{2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2 \xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})}}{2\pi \left(\det\left[\nabla^2 \xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}}} d\eta_l d\eta_{l+k} \\
& + \frac{1}{2} \int\int (\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2 a(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) \frac{e^{-\frac{1}{2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})^T \nabla^2 \xi(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})}}{2\pi \left(\det\left[\nabla^2 \xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}}} d\eta_l d\eta_{l+k}. \qquad (4.13)
\end{aligned}$$

The first line in Eq. (4.13) integrates to $a(\hat{\boldsymbol{\eta}})$. The second line can be seen as $\nabla a(\hat{\boldsymbol{\eta}})\mathrm{E}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})$, where $\boldsymbol{\eta}$ is a bivariate normally distributed random vector with mean $\hat{\boldsymbol{\eta}}$; therefore, this integrates to 0. For the integral in the third line, use the substitution $\mathbf{Z} = \left(\nabla^2 a(\hat{\boldsymbol{\eta}})\right)^{\frac{1}{2}} (\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})$, with Eq. (4.13) becoming

$$\begin{aligned}
\mathrm{I}_2 &\approx a(\hat{\boldsymbol{\eta}}) + \frac{1}{2} \int\int \mathbf{z}^T \mathbf{z} \frac{e^{-\frac{1}{2}\mathbf{z}^T \left(\nabla^2 a(\hat{\boldsymbol{\eta}})\right)^{-\frac{1}{2}} \nabla^2 \xi(\hat{\boldsymbol{\eta}}) \left(\nabla^2 a(\hat{\boldsymbol{\eta}})\right)^{-\frac{1}{2}} \mathbf{z}}}{2\pi \left(\det\left[\nabla^2 \xi(\hat{\boldsymbol{\eta}})\right]\right)^{-\frac{1}{2}} \left(\det\left[\nabla^2 a(\hat{\boldsymbol{\eta}})\right]\right)^{\frac{1}{2}}} d\mathbf{z} \\
&= a(\hat{\boldsymbol{\eta}}) + \frac{1}{2} \mathrm{tr}\left\{ \left[\nabla^2 \xi\left(\hat{\boldsymbol{\eta}}\right)\right]^{-1} \nabla^2 a\left(\hat{\boldsymbol{\eta}}\right) \right\}.
\end{aligned}$$

The integral in the first line of the above equation is the expectation $\mathrm{E}\left(\mathbf{Z}^T\mathbf{Z}\right)$, where $\mathbf{Z}$ is bivariate normal random vector with mean 0 and variance $\left(\nabla^2 a(\hat{\boldsymbol{\eta}})\right)^{\frac{1}{2}} \nabla^2 \xi(\hat{\boldsymbol{\eta}})^{-1} \left(\nabla^2 a(\hat{\boldsymbol{\eta}})\right)^{\frac{1}{2}}$. This leads immediately to the expression in the second line.

4.6 Simulation Study Results

A simulation study was performed to compare the performance of the composite likelihood, with and without adjustment, and the SM method, in which the full likelihood is used. The design of this simulation study is the same to the one described in Section 3.4, which we will briefly summarize. We studied two orders of the latent process, AR(2) and AR(3). For each order, we studied three different scenarios for the correlation structure (weak, moderate and strong), each with two values for σ_V. This was done for three different sample sizes, a small size $n = 20$, and two large sample sizes, $n = 365$ and $n = 1095$. Therefore, we have 36 different

cases; however, we noticed that there is no pronounced difference in the results between $AR(2)$ or $AR(3)$. Therefore, only the results of $AR(2)$ with larger σ_V will be presented here. The complete results can be found in Tables B.1 to B.18 in appendix B. We have also noticed that there is no big improvement on the results of the composite likelihood after $m = 3$ for weak and moderate serial correlation and after $m = 5$ for strong serial correlation.

The results of the simulation study assuming the $AR(2)$ latent process is summarized in Figures 4.1 to 4.6. From the results, we can conclude the following:

First, the composite posterior without adjustment produce falsely lower RMSE for $\boldsymbol{\beta}$ and σ_V than the full posterior in all cases. This conclusion was not true for $\boldsymbol{\phi}$, where the composite posterior produces higher RMSE than the full posterior.

Second, the magnitude adjustment always offers higher RMSE for the parameter estimates than does the composite posterior without adjustment. In contrast, both the curvature and OFS adjustment methods adjust the RMSE produced by the composite posterior, and make it closer to the RMSE produced by the full posterior.

Third, for decent and large sample sizes, $n = 365$ and $n = 1095$, the curvature and the OFS adjustments performed almost the same in terms of bias and RMSE of parameter estimates. Both methods outperform magnitude adjustment by having RMSE of all parameters closer to the full posterior. The only exception is β_0 in case of weak and moderate correlation structure, where magnitude adjustment offers RMSE closer to that of the full posterior than do the curvature and OFS adjustment methods. In terms of bias, all adjustment methods tend to have a similar bias for $\boldsymbol{\beta}$, and it is very close to the bias of the full posterior. The only exception is for β_0 in the case of the strong correlation structure, where the bias for the adjusted composite likelihood is smaller than the bias for the full likelihood. When the sample size is decent and the value of σ_V is small, the curvature and OFS adjustments may perform differently in terms of the RMSE of $\boldsymbol{\phi}$. In such a case, the OFS tend to have RMSE of $\boldsymbol{\phi}$ closer to the full likelihood than does the curvature for weak and moderate serial correlation, while the opposite is true for the stronger correlation.

Finally, for the small sample size $n = 20$ and weak and moderate correlations, curvature

adjustment offers the RMSE of $\boldsymbol{\beta}$ closest to the full likelihood than other adjustment methods, followed by the OFS. On the other hand, the magnitude adjustment has the RMSE of $\boldsymbol{\phi}$ closest to the full likelihood than other adjustments. The curvature adjustment also offers RMSE for σ_V closest to the full in the case of higher σ_V.

In conclusion, the curvature adjustment method tends to perform at least as well as the OFS adjustment method in terms of bias and RMSE for small sample size. However, both the curvature and OFS adjustment methods performed almost the same for decent and large sample sizes. In almost all cases, the curvature and OFS adjustment methods outperform the magnitude adjustment method. In terms of the computational time, the OFS adjustment method took the minimum time, followed by both the curvature and the magnitude adjustment methods. The computational time for the OFS adjustment method with $m = 3$ was very close to that of the SM method, while the curvature and the magnitude adjustment methods needed almost twice the time. Therefore, we recommend the use of curvature adjustment for small sample size and OFS for decent and larger sample sizes.

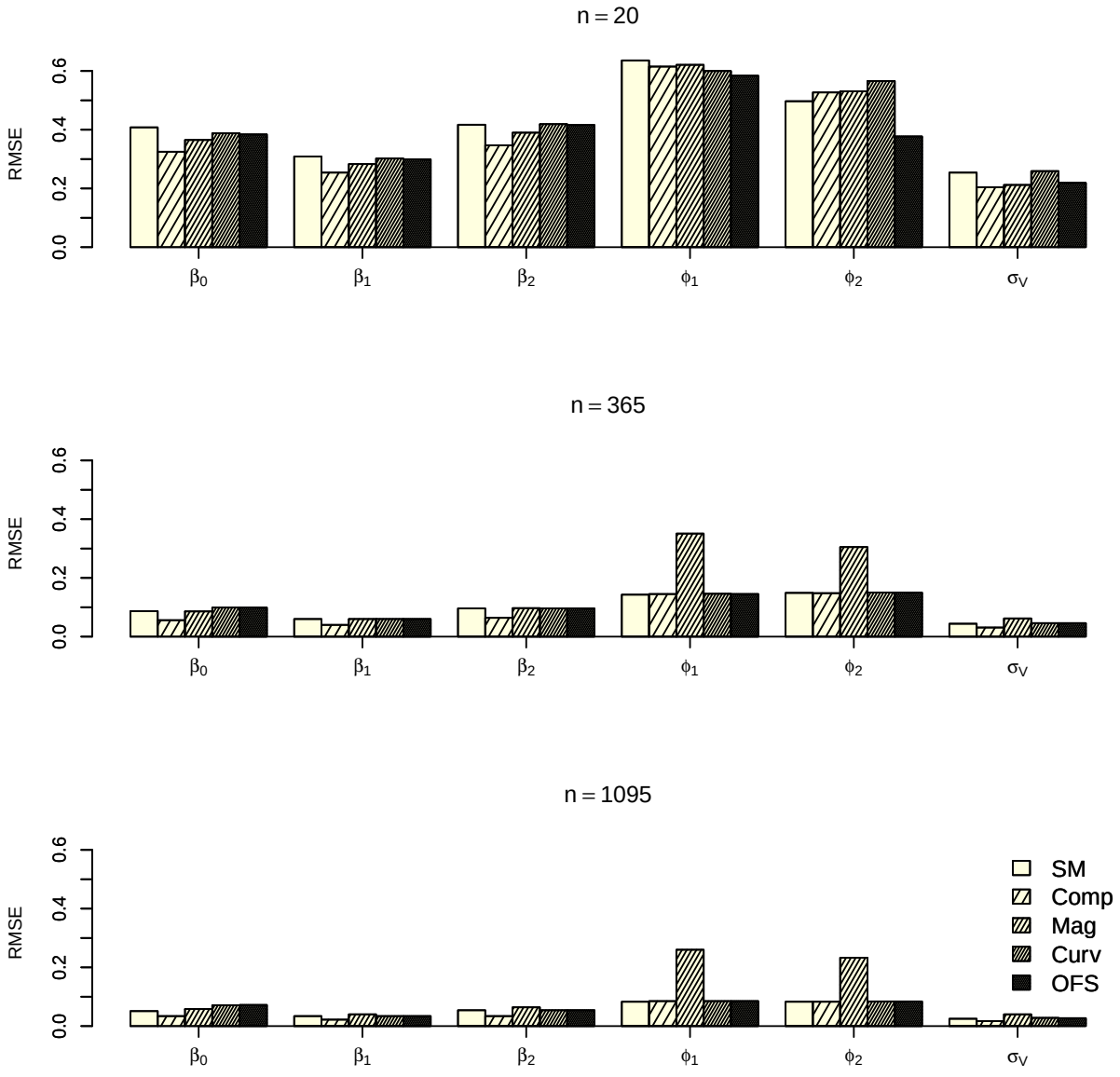

Figure 4.1: RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.06)$ and $\sigma_V = 0.4$.

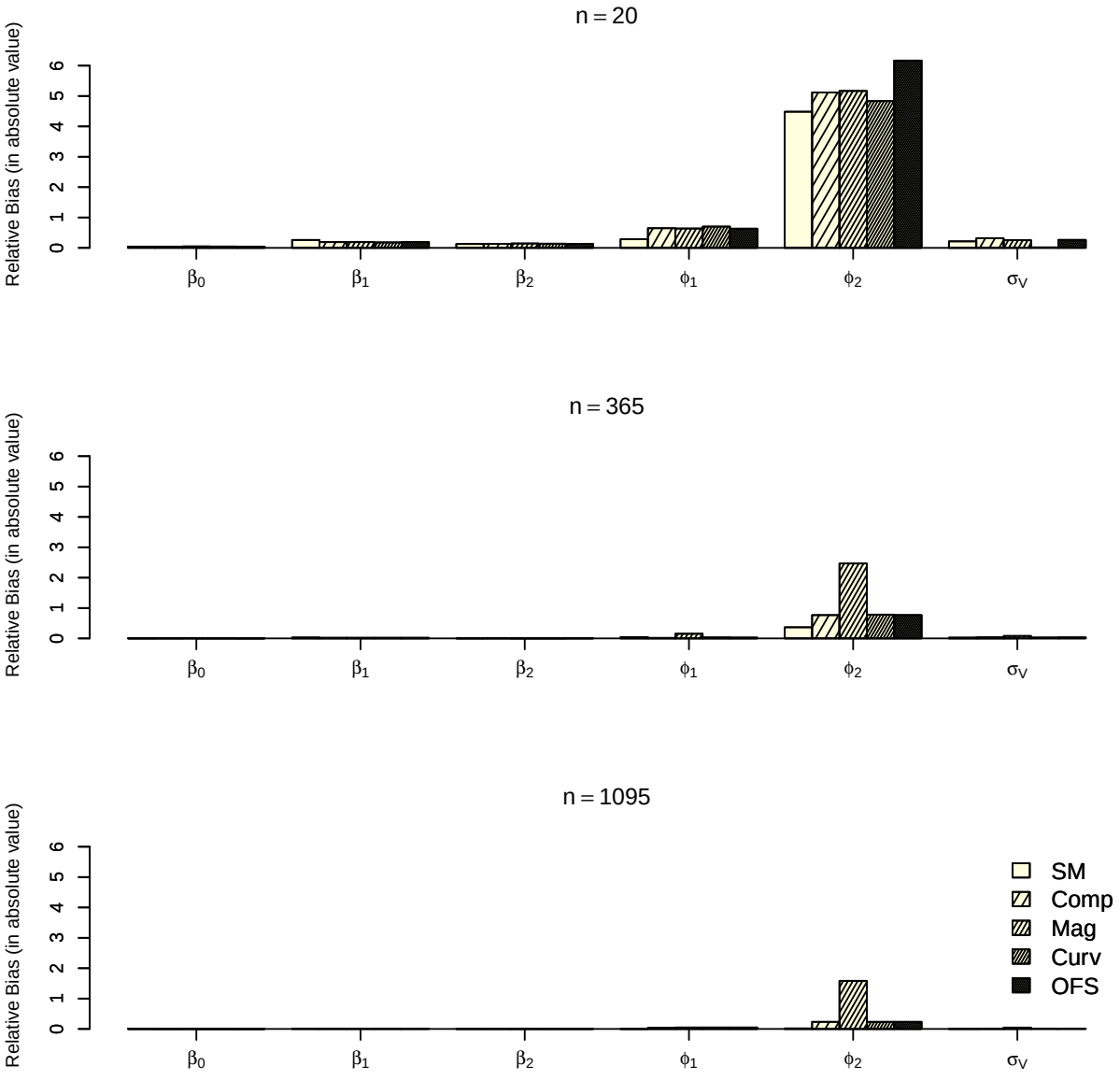

Figure 4.2: Relative bias (in absolute value) of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.06)$ and $\sigma_V = 0.4$.

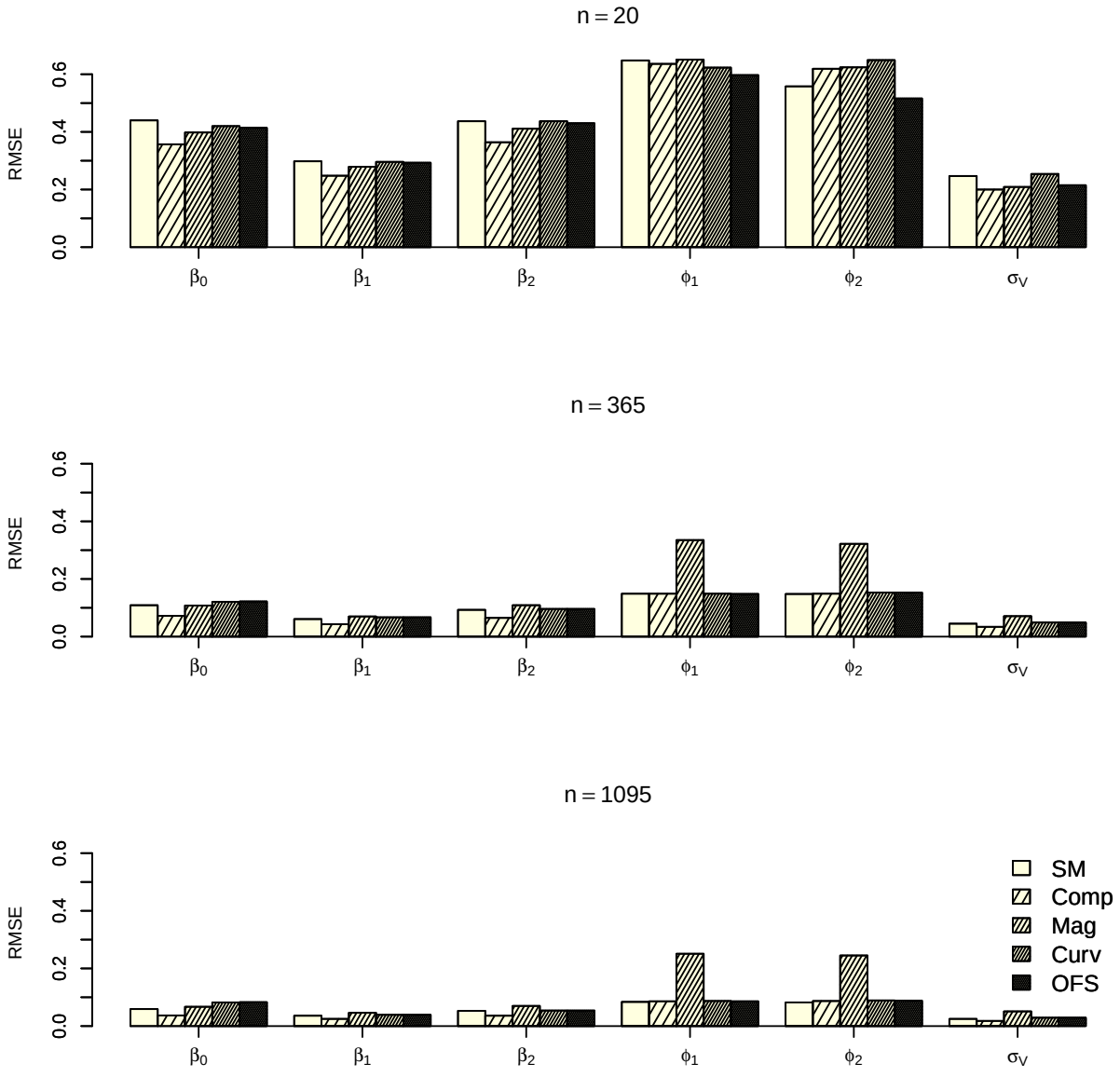

Figure 4.3: RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $\sigma_V = 0.4$.

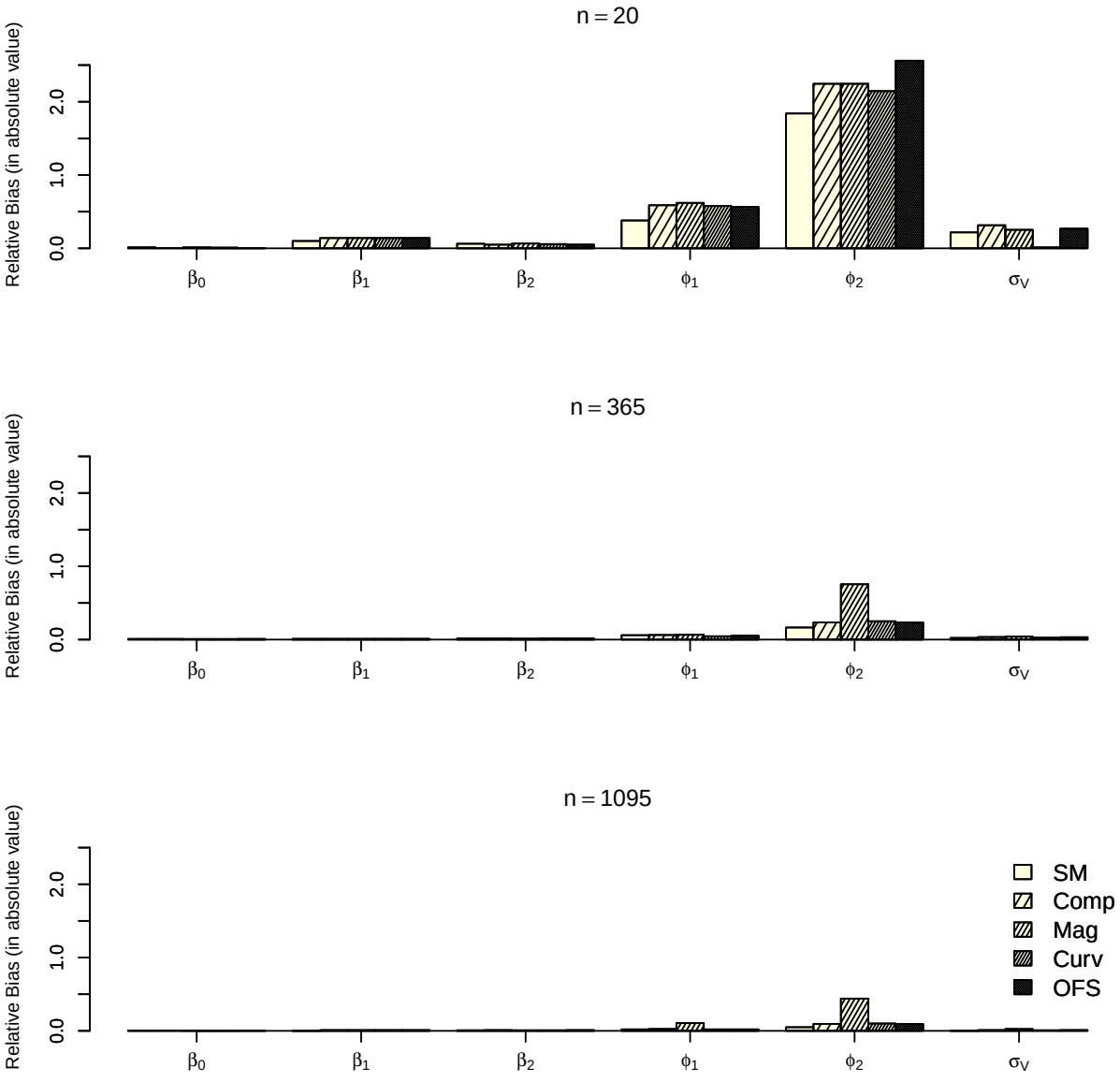

Figure 4.4: Relative bias (in absolute value) of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $\sigma_V = 0.4$.

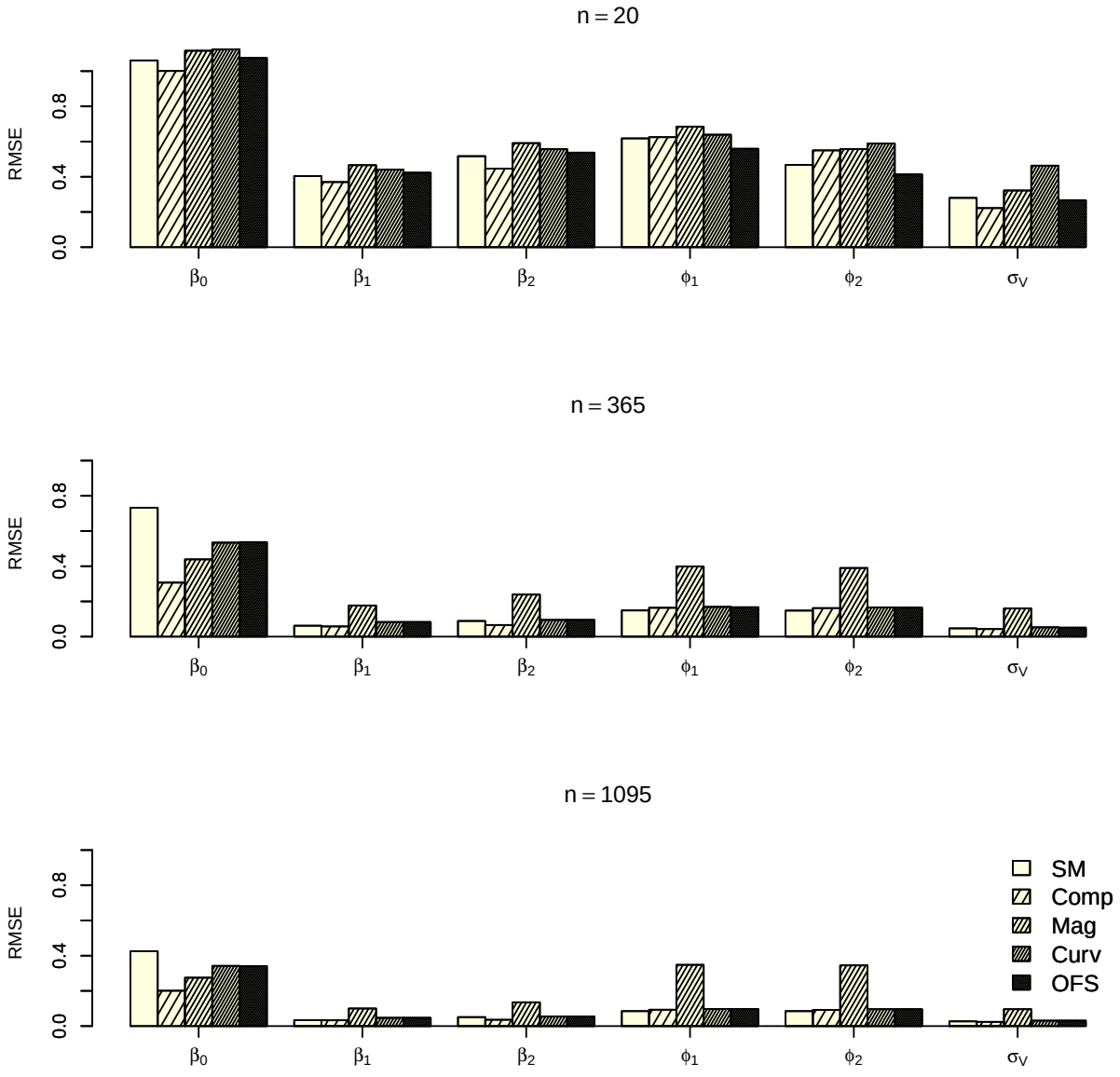

Figure 4.5: RMSE of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $\sigma_V = 0.4$.

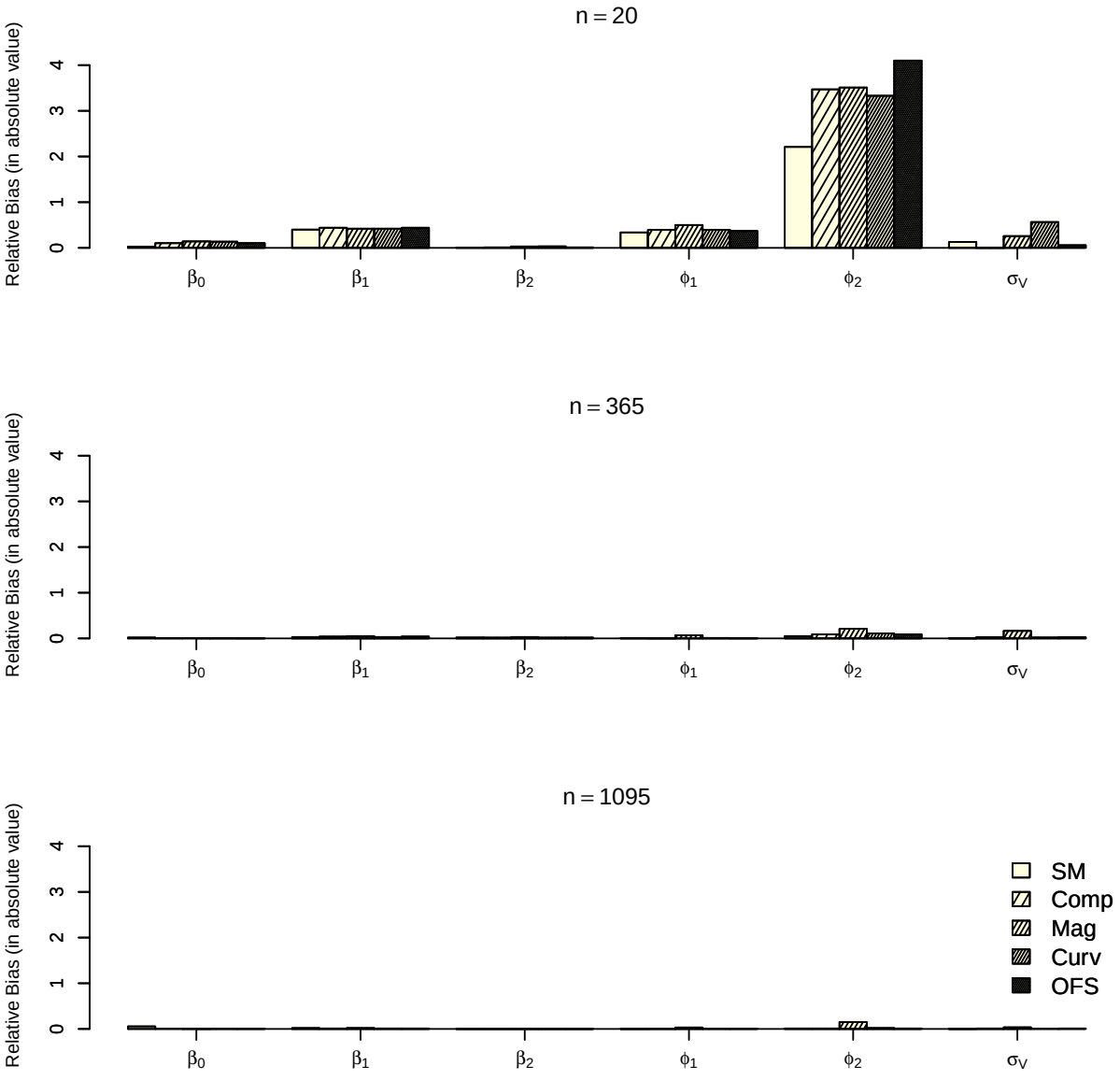

Figure 4.6: Relative Bias (in absolute value) of parameter estimates for SM and composite likelihood methods with the $AR(2)$ latent process; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $\sigma_V = 0.4$.

Chapter 5
Analysis of Time Series of Car Crash Data

5.1 Data Description

The aim of this chapter is to illustrate the application of methods discussed in previous chapters via a real data example. This data example consists of a time series of daily crash counts on a major road near Schiphol, the largest airport in the Netherlands, in 2001. This data set is from Sermaidis (2006). The purpose of the application is to study the effects of weather conditions on the number of daily car crashes. In particular, we are interested in what covariates are important to the occurrence of car accidents.

The available data consist of crash counts as a response variable and some weather covariates that are available on a daily basis for the year of 2001. Therefore, we have $n = 365$ observations, which do not have missing values. The available covariates are: (1) WD=cosine of twice the mean of the prevailing wind direction in degrees. The underlying degrees have the following values: North=360, South=180, West=270, and calm/variable=0, (2) IWD=indicator of weekday (one for weekdays and zero for weekends), (3) RD=mean hourly radiations in Joule/cm^2 as a measure of the intensity of the sun or how much the sun reaches the earth, and (4) PD=average number of periods per hour (each period is 0.1 hour units) that had precipitation. This was found by dividing the day into 240 time periods, each of length 0.1 hour, and then dividing the number of periods that experienced precipitation by 24. More details about the variable descriptions can be found in Brijs et al. (2008) and Sermaidis (2006).

We will discuss the analysis of this data set in the Bayesian paradigm using all methods discussed in the thesis, namely SM, PGS and Bayesian composite likelihood methods. For the Bayesian composite likelihood adjustment methods, only the OFS adjustment method will be used here, because the findings from the simulation studies in Chapter 4 suggest OFS would be favored when sample size is decent, which is the case in this data. The posterior summaries of parameters are reported in Table 5.1 for all aforementioned methods.

This car accident data has been analyzed before by both Brijs et al. (2008) and Ng et al. (2011). The former assumed a Poisson integer valued autoregressive model (INAR) for the data. In contrast, the latter analyzed the data assuming a Poisson parameter-driven model. The advantage of the Poisson parameter-driven model over the Poisson INAR is that it accounts for overdispersion in the data, while the Poisson INAR model does not, as it assumes a Poisson marginal distribution. Two common properties between the two models that they both are frequentist methods and can account for serial correlation in the data. Ng et al. (2011), in their analysis of this dataset, compare the performance of composite likelihood and Zeger's methods with the approximate maximum likelihood via MCMC.

5.2 Analysis and Results

Before we analyze the data using methods discussed in this thesis, we will use the Deviance Information Criterion (DIC) method by Spiegelhalter et al. (2002) to select the 'best' model. The calculation of DIC requires the evaluation of the likelihood function, which, in our case, involves high dimensional integral. Several constructions of the DIC has been proposed in literature to address this issue; see, for example, Celeux et al. (2006), Mason et al. (2012), and Li et al. (2012). In this thesis, we will use the built-in method for DIC calculation in the rjags package, which was suggested by Plummer (2002).

All models compared by DIC will assume a Poisson parameter-driven model for counts, but with different orders of the $AR(p)$ process and different combinations of the covariates. In the first step, we will include all covariates in the model and select p that gives the smallest

value of DIC. In the second step, we assumed the $AR(p)$ process selected in the first step, and then compared models with different combinations of the covariates. A backward elimination method was followed to select the covariates in the second step. The model that offered the smallest DIC value was selected. The selected model was a Poisson parameter-driven model with $AR(2)$ latent process and three covariates: WD, IWD and PD.

Given the selected model, we now can analyze the data using the methods mentioned in Section 5.1. Table 5.1 shows the posterior means, standard deviations and credible intervals of all parameters using SM, PGS, composite likelihood, and OFS adjustment methods. The results agree with the simulation studies presented in Chapter 3 and Chapter 4. First, the PGS offers very close results to the SM for all parameters. Second, the Bayesian composite likelihood method without adjustment produced very close posterior means to the SM, but the posterior standard deviations are noticeably lower, especially for the $\boldsymbol{\beta}$'s and σ_V, which in turn led to narrower CI's. With the OFS adjustment method, we notice that the means of the parameters did not change and remain very close to the SM. Regarding the standard deviations, OFS adjustment has altered the standard deviations of all parameters and made them closer to the results of the SM method. Third, we noticed that $m = 1$ is good enough for the composite likelihood to produce estimates for $\boldsymbol{\beta}$ and σ_V that are close to results of the SM method, while a larger value, $m = 3$, is required to have close results for $\boldsymbol{\phi}$. No big changes in the composite likelihood results were observed for $m > 3$.

Based on the results from the SM method, the values of ϕ_1 and ϕ_2 reveal a moderate serial correlation structure ($\rho_1 = 0.50$, $\rho_2 = 0.41$). The effect of weather covariates on the number of car crashes can be interpreted as follows: first, WD does not appear to be significant; the 95% C.I is (-0.123,0.017), and, second, both the IWD and PD are positively affecting the number of car crashes. The coefficient of IWD is $\beta_2 = 0.473$ with a 95%C.I (0.365,0.581), which means that weekdays increase the expected number of car crashes by 2 crashes while keeping other covariates unchanged. The coefficient of PD is $\beta_3 = 0.174$ with a 95% C.I (0.139,0.208), that is, the expected number of car crashes would increase by one crash as PD increases by one unit, while keeping other covariates unchanged.

Table 5.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters using SM, PGS, composite likelihood, and OFS methods for Car crash data with an $AR(2)$ latent process and three covariates (WD, IWD, PD).

Par.		SM	PGS	Comp.	OFS
Intercept	Mean(SD)	1.586(0.064)	1.589(0.064)	1.597(0.022)	1.597(0.068)
	95%CI	(1.452, 1.710)	(1.459, 1.712)	(1.553, 1.641)	(1.464, 1.731)
WD	Mean(SD)	−0.054(0.036)	−0.055(0.036)	−0.055(0.015)	−0.055(0.037)
	95%CI	(−0.123, 0.017)	(−0.128, 0.014)	(−0.084, −0.027)	(−0.129, 0.017)
IWD	Mean(SD)	0.473(0.055)	0.472(0.054)	0.468(0.023)	0.468(0.053)
	95%CI	(0.365, 0.581)	(0.363, 0.580)	(0.424, 0.512)	(0.367, 0.570)
PD	Mean(SD)	0.174(0.017)	0.174(0.017)	0.177(0.007)	0.177(0.019)
	95%CI	(0.139, 0.208)	(0.142, 0.211)	(0.163, 0.191)	(0.141, 0.213)
ϕ_1	Mean(SD)	0.380(0.144)	0.389(0.140)	0.344(0.155)	0.337(0.140)
	95%CI	(0.153, 0.717)	(0.154, 0.697)	(0.124, 0.725)	(0.129, 0.669)
ϕ_2	Mean(SD)	0.224(0.140)	0.225(0.140)	0.248(0.179)	0.250(0.159)
	95%CI	(−0.093, 0.469)	(−0.069, 0.486)	(−0.137, 0.538)	(−0.093, 0.506)
σ_V	Mean(SD)	0.274(0.031)	0.273(0.031)	0.270(0.016)	0.271(0.034)
	95%CI	(0.210.0.335)	(0.213, 0.334)	(0.234, 0.300)	(0.201, 0.336)

5.3 Model Adequacy Check

Using DIC, we can select the 'best' model among the proposed candidates; however, there still remains the problem of deciding whether the selected model is adequate for the data at hand. To answer this question, methods for model adequacy checking have to be used. Limited work can be found on how to assess model adequacy for SSM. Altman (2004) proposed a graphical technique for assessing the goodness of fit for a stationary hidden Markov model (HMM). The author proposed to plot the estimated distribution function for the observed data against the empirical distribution function. If the plot forms a straight line with 45° through the origin, then this suggest that the proposed model is valid. A plot of multidimensional distribution functions can also be considered. However, this method is valid if the data is assumed to be identically distributed, which is not the case of our proposed model. Other goodness of fit tests for dependent but identically distributed observations can be found in Ignaccolo (2004). Goodness of fit tests for the case with the Poisson assumption and not identically, but independently, distributed observations can be found in McCullagh (1986) and Collings and

Margolin (1985). In the aforementioned goodness of fit methods, only McCullagh's method can handle covariates in the model; however, it assumes independent observations.

In this chapter, we will use the 'ordinary pseudo-residuals' and the 'forecast pseudo-residuals' methods proposed by Macdonald and Zucchini (2009) in the context of HMM analysis. These two methods can be used for dependent observations and can handle covariates in the model. They are based on the conditional distribution of the current observation given other observations. The difference between ordinary and forecast pseudo-residuals is that the former is based on the conditional distribution given all other observations, while the latter is based on the conditional distribution given all preceding observations. The ordinary pseudo-residuals, for discrete observations, are the line segments $\left[z_t^-; z_t^+\right]$, where

$$z_t^- = \Phi^{-1}\left(u_t^-\right),$$

and

$$z_t^+ = \Phi^{-1}\left(u_t^+\right),$$

where Φ^{-1} is the inverse distribution function of the standard normal distribution, and

$$u_t^- = P(Y_t < y_t | Y_{(t)} = y_{(t)}), \tag{5.1}$$

and

$$u_t^+ = P(Y_t \leq y_t | Y_{(t)} = y_{(t)}), \tag{5.2}$$

where $y_{(t)}$ is the vector of all observations except y_t, i.e.

$$y_{(t)} = \{y_1, y_2, \ldots, y_{t-1}, y_{t+1}, \ldots, y_n\}.$$

If the assumed model is true, then the ordinary and the forecast pseudo-residuals follow a standard normal distribution. Therefore, to check the model adequacy, we compute the ordinary and the forecast pseudo-residuals and then plot the normal qq-plots for each of them. If the plots form approximately a straight 45° line, then we conclude that the assumed model is valid.

The probabilities, u_t^- and u_t^+, given in Eq. (5.1) and Eq. (5.2), are defined in the frequentist paradigm. We borrow the idea of posterior predictive distribution from Gelman et al. (1996) and redefine them as follows. Let Y_t^{rep}, $t = 1, \ldots, n$, be the replicated data that could have been observed at time t if the experiment that produced Y were replicated with the same model M, the same values of the covariates and the same value of the parameter $\boldsymbol{\theta}$.

$$u_t^- = P(Y_t^{rep} < y_t | M, Y_{(t)}^{rep} = y_{(t)}, Y_{1:n} = y_{1:n}), \tag{5.3}$$

and

$$u_t^+ = P(Y_t^{rep} \leq y_t | M, Y_{(t)}^{rep} = y_{(t)}, Y_{1:n} = y_{1:n}), \tag{5.4}$$

In the following, we will discuss the evaluation of the ordinary pseudo-residuals based on the probabilities in Eq. (5.3) and Eq. (5.4). The evaluation of the forecast pseudo-residuals will be conducted in a similar way, except that u_t^+ and u_t^- are defined as follows:

$$u_t^- = P(Y_t^{rep} < y_t | M, Y_{1:(t-1)}^{rep} = y_{1:(t-1)}, Y_{1:n} = y_{1:n}),$$

and

$$u_t^+ = P(Y_t^{rep} \leq y_t | M, Y_{1:(t-1)}^{rep} = y_{1:(t-1)}, Y_{1:n} = y_{1:n}).$$

The probabilities in Eq. (5.3) and Eq. (5.4) can be evaluated using MCMC sampling. First, let us decompose the conditional density of $Y_t^{rep} | Y_{(t)}^{rep} = y_{(t)}, Y_{1:n} = y_{1:n}$ as follows. For simplicity of notation, we will drop the M from the equations below.

$$
\begin{aligned}
&f_{Y_t^{rep}|Y_{(t)}^{rep}=y_{(t)},Y_{1:n}=y_{1:n}}\left(y_t|y_{(t)},y_{1:n}\right)= \\
&\frac{\Pr\left(Y_1^{rep}=y_1,\ldots,Y_n^{rep}=y_n|y_{1:n}\right)}{\Pr\left(Y_1^{rep}=y_1,\ldots,Y_{t-1}^{rep}=y_{t-1},Y_{t+1}^{rep}=y_{t+1},\ldots,Y_n^{rep}=y_n|y_{1:n}\right)} \\
&=\frac{\int\cdots\int f_{Y_{1:n}^{rep},\eta_{1:n},\boldsymbol{\theta}|Y_{1:n}=y_{1:n}}\left(y_{1:n},\eta_{1:n},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}}{\int\cdots\int f_{Y_{(t)}^{rep},\eta_{(t)},\boldsymbol{\theta}|Y_{1:n}=y_{1:n}}\left(y_{(t)},\boldsymbol{\eta},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}} \\
&=\frac{\int\cdots\int\left\{\prod_{j=1}^{n}g(y_j|\eta_j,\boldsymbol{\theta})\right\}\pi\left(\eta_{1:n},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}}{\int\cdots\int\left\{\prod_{j=1;j\neq t}^{n}g(y_j|\eta_j,\boldsymbol{\theta})\right\}\pi\left(\eta_{1:n},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}},
\end{aligned}
\tag{5.5}
$$

where g is the Poisson probability mass function and π is the joint posterior density of $\{\boldsymbol{\eta},\boldsymbol{\theta}\}$. The integrals in Eq. (5.5) can be approximated by the empirical mean of $\left\{\prod_{j=1}^{n}g(y_j|\eta_j,\boldsymbol{\theta})\right\}$ and $\left\{\prod_{j=1;j\neq t}^{n}g(y_j|\eta_j,\boldsymbol{\theta})\right\}$, respectively, as follows. Suppose that we have J MCMC samples of $\{\boldsymbol{\eta},\boldsymbol{\theta}\}$, i.e., $\left\{\boldsymbol{\eta}^{(1)},\ldots,\boldsymbol{\eta}^{(J)}\right\}$ and $\left\{\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(J)}\right\}$, then the empirical means, M_1 and M_2, respectively, are:

$$
\mathrm{M}_1=\frac{1}{J}\sum_{j=1}^{J}\prod_{t=1}^{n}g(y_t|\eta_t^{(j)},\boldsymbol{\theta}^{(j)}),
$$

and

$$
\mathrm{M}_2=\frac{1}{J}\sum_{j=1}^{J}\prod_{k=1;k\neq t}^{n}g(y_k|\eta_k^{(j)},\boldsymbol{\theta}^{(j)}).
$$

Therefore, the the probability mass function $f_{Y_t^{rep}|Y_{(t)}^{rep}=y_{(t)}}\left(y_t|y_{(t)}\right)$ can be approximated by

$$
\frac{\mathrm{M}_1}{\mathrm{M}_2}.
$$

Please note that the MCMC samples can be obtained using the SM method discussed in Chapter 2.

In a similar way, the density of $Y_t^{rep}|Y_{1:(t-1)}^{rep}=y_{1:(t-1)},Y_{1:n}=y_{1:n}$ can be written as follows

$$f_{Y_t^{rep}|Y_{1:(t-1)}^{rep}=y_{(t)},Y_{1:n}=y_{1:n}}\left(y_t|y_{1:(t-1)},y_{1:n}\right) =$$
$$= \frac{\int\cdots\int\left\{\prod_{j=1}^{t}g(y_j|\eta_j,\boldsymbol{\theta})\right\}\pi\left(\eta_{1:n},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}}{\int\cdots\int\left\{\prod_{j=1}^{t-1}g(y_j|\eta_j,\boldsymbol{\theta})\right\}\pi\left(\eta_{1:n},\boldsymbol{\theta}|y_{1:n}\right)d\boldsymbol{\eta}d\boldsymbol{\theta}}, \tag{5.6}$$

The normal qq-plots of the ordinary pseudo-residuals and forecast pseudo-residuals are given in Figures 5.1. Based on the two plots, it can be seen that there is no serious deviation from normality. Therefore, we would accept the assumed model.

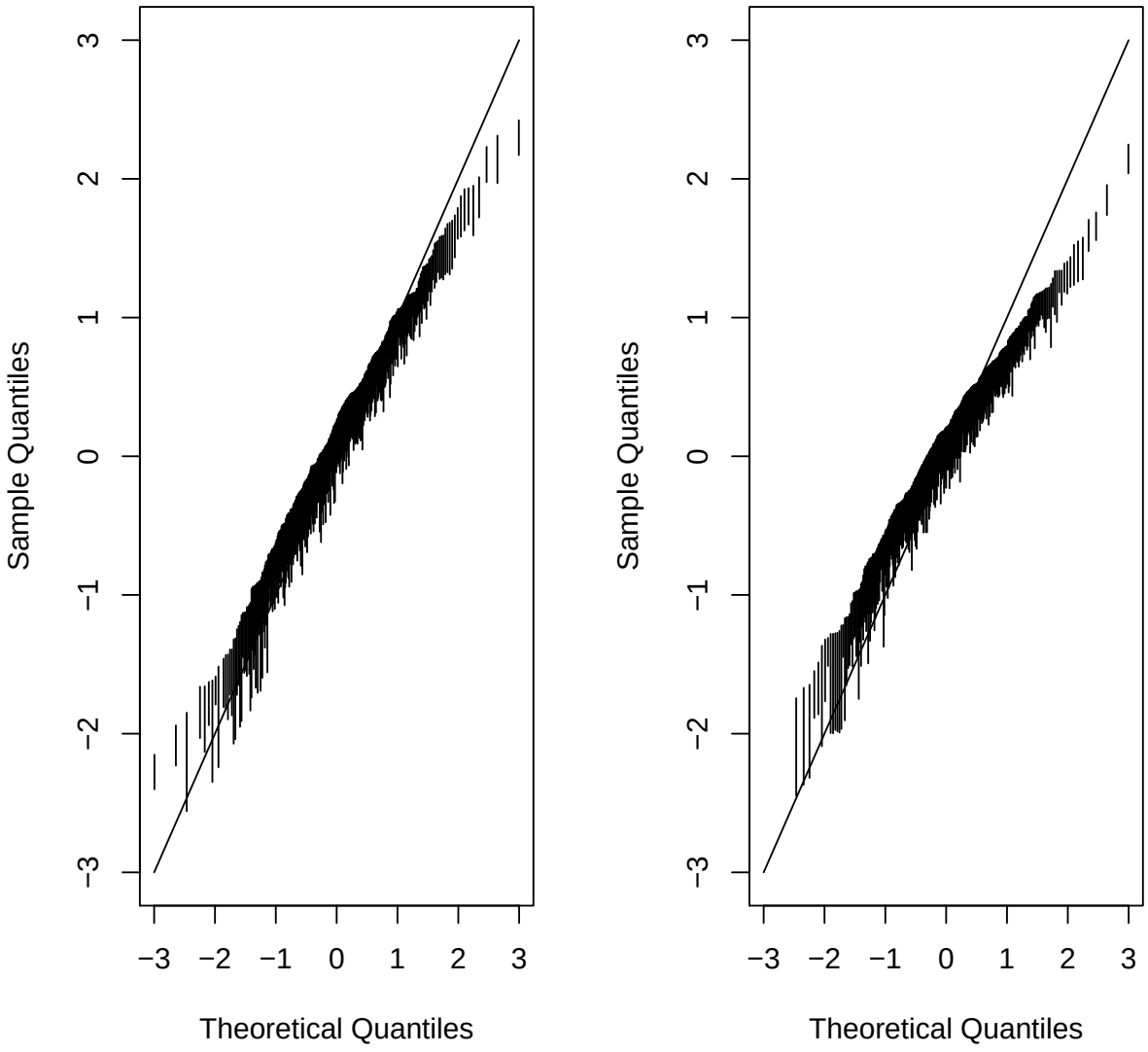

Figure 5.1: Normal qq-plot of ordinary pseudo-residuals (left) and forecast pseudo-residuals (right) for car crashes data.

Chapter 6

Discussion and Future Plan

6.1 Main Results

This thesis focuses on Bayesian estimation of the Poisson parameter-driven model. The Poisson parameter-driven model involves an autoregressive latent process to account for temporal dependence between observations. The inference of this model is complicated, as the likelihood function of the observed data involve high dimensional integral over the joint distribution of the latent process. In this thesis, we concentrate on MCMC methods. First, we develop a new single-move (SM) MCMC method to sample the latent process one by one using an accept-reject method. Other variables were sampled either directly from their full conditional posteriors or using the Metropolis-Hastings algorithm. Second, with a hope of improving the mixing performance of the MCMC algorithm, we apply the particle Gibbs sampler (PGS) method in current model settings to sample the latent process as a group, and compare it with the SM method. In the PGS method, the conditional SMC methods are used to recursively design a high dimensional proposal to sample the latent process within the MCMC method. Other parameters are sampled in the same way as in the SM method. Third, we consider Bayesian composite likelihood methods to analyze the Poisson parameter-driven model, and compare their performance with the SM method. The use of composite likelihood method replaces the high dimensional integral in the likelihood function by low dimensional ones, which simplifies the integration of the latent process. Adjusting the composite likelihood is necessary to provide valid inference. Three different adjustment methods for the composite likelihood are discussed and compared. The latter two proposed methods

(PGS and Composite likelihood) were compared with the SM method through simulation studies as well as via a real data example.

Results from the simulation studies show that the PGS is almost similar to the SM method in terms of bias and RMSE of the parameter estimates. However, the PGS is more time consuming, which primarily due to the resampling step that has to be performed at each time point. The composite likelihood approach needed some adjustments to provide results very close to the SM method. Three adjustment methods were discussed – magnitude, curvature and OFS adjustment methods. The recommended adjustment methods were the curvature method for small sample size and the OFS for large sample size. The time needed for the composite likelihood with adjustment varies from very similar (in case of OFS method) to more than twice (in case of curvature and magnitude methods) the time needed for the SM method. For example, with an Intel Xeon 2.67 GHz processor, a sample size $n = 365$ and 28000 iterations, the computational times in the C language are 15, 18, 35, and 40 seconds for SM, OFS, magnitude and curvature methods respectively. The findings from the real data example agree with the simulation study. First, the posterior means and standard deviations of the parameters using the PGS method were very close to those using the SM method. Second, the composite likelihood methods with OFS adjustment provide very close results to the SM method, while the unadjusted composite likelihood method provide a smaller standard deviation of the parameters, especially for $\boldsymbol{\beta}$ and σ_V.

6.2 Future Work

Below we outline a number of important future directions that can enhance the current work.

- To our best knowledge, there is no formal goodness of fit test available in the context of a Poisson parameter driven model with a latent AR process. My interest is to develop a goodness of fit test to accommodate correlated and non-identically distributed Poisson count data.

- In the frequentist paradigm, inference based on the composite likelihood is known to be

robust to model misspecification. Therefore, my interest is to study the accuracy and efficiency of the composite Bayes estimate (based on adjusted composite likelihood) and compare it to the Bayes estimate (based on full likelihood) when the full likelihood is misspecified.

- In the current model, we consider the autoregressive process to model the dependence between observations. In practice, different types of time series models may be considered, such as the autoregressive moving average process (ARMA).

- The focus of this thesis is the Poisson parameter-driven model. In this model, it is possible to observe zero counts. However, in some applications, observing a count of zero is impossible. For example, in ecological studies, the researchers may be interested in the group size of an animal under study. Such data is called zero-truncated data. In other applications, the data may contain a large number of zero counts and, hence, is called zero-inflated data. One example is in epidemiological studies, when the disease of interest exhibits low case counts during regular periods and, hence, a high frequency of zeros could be observed, but where counts can also be large during pandemic periods. My interest is to develop Bayesian approaches for estimating zero-inflated and zero-truncated Poisson parameter-driven models.

References

R.M. Altman. Assessing the goodness-of-fit of hidden Markov models. *Biometrics*, 60:444–450, 2004.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods (with discussion). *JRSS(B)*, 72:269–342, 2010.

T. Apanasovich, D. Ruppert, J. Lupton, N. Popovic, and R. Carroll. Aberrant crypt foci and semiparametric modeling of correlated binary data. *Biometrics*, 64(2):490–500, 2008.

A. Azzalini. Maximum likelihood estimation of order m for stationary stochastic processes. *Biometrika*, 70:381–387, 1983.

O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics (with discussion). *JRSS(B)*, 63:167–241, 2001.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *JRSS(B)*, 36(2): 192–236, 1974.

T. Brijs, D. Karlis, and G. Wets. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis and Prevention*, 40:1180–1190, 2008.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1991.

A. C. Cameron and P. K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, UK, 1998.

G. Celeux, F. Forbes, C. P. Robert, and Titterington D. M. Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–674, 2006.

K. S. Chan and J. Ledolter. Monte Carlo estimation for time series models involving counts. *JASA*, 90:242–252, 1995.

R. E. Chandler and S. Bate. Inference for clustered data using the independence log-likelihood. *Biometrika*, 94:167–183, 2007.

V. Chernozhukov and H. Hong. An MCMC approach to classical estimation. *Journal Of Econometrics*, 115(2):293–346, 2003.

B. J. Collings and B. H. Margolin. Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *JASA*, 80(390):411–418, 1985.

D. Cox and N. Reid. A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737, 2004.

D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

D. R. Cox. Statistical analysis of time series: some recent developments. *Scand. J Statist.*, 8:93–115, 1981.

R. Davis and C.Y. Yau. Comments on pairwise likelihood in time series models. *Statistica Sinica*, 21:255–277, 2011.

P. Del Moral, A. Doucet, and Jasra A. Sequential Monte carlo samplers. *JRSS(B)*, 68(3): 411–436, 2006.

R. Douc, O. Cappe, and E. Moulines. Comparison of resampling schemes for particle filtering. *In 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2005.

J. Durbin and S. J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *JRSS(B)*, 62:3–56, 2000.

T. S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall, London, UK, 1996.

N. Friel. Bayesian inference for Gibbs random fields using composite likelihoods. *Proceedings of the 2012 Winter Simulation Conference*, 2012.

S. Fruhwirth-Schnatter and H. Wagner. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93 (4):827–841, 2006.

S. Fruhwirth-Schnatter, R. Fruhwirth, L. Held, and H. Rue. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19: 479–492, 2009.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

A. Gelman, X. Meng, and Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–807, 1996.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

H. Geys, G. Molenberghs, and L. Ryan. Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *JASA*, 94:734–745, 1999.

W. R. Gilks and C. Berzuini. Following a moving target-Monte carlo inference for dynamic Bayesian models. *JRSS(B)*, 63:127–146, 2001.

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140:107–113, 1993.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

P. Heagerty and S. Lele. A composite likelihood approach to binary spatial data. *JASA*, 93: 1099–1111, 1998.

R. Ignaccolo. Goodness of fit tests for dependent data. *Nonparametric Statistics*, 16(1-2): 19–38, 2004.

Z. Jin. *Aspects of Composite Likelihood Inference*. PhD thesis, University of Toronto, 2010.

H. Joe. Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, 52:5066–5074, 2008.

M. C. Jones. Randomly choosing parameters for the stationarity and invertibility region of autoregressive-moving average models. *Appl. Statist.*, 36(2):134–138, 1987.

R. C. Jung, M. Kukuk, and R. Liesenfeld. Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis*, 51:2350–2364, 2006.

J. Kent. Robust properties of likelihood ratio tests. *Biometrika*, 69:19–27, 1982.

A. Kuk and Y. Cheng. The Monte carlo Newton-Raphson algorithm. *J. of Statistical Computation and Simulation*, 59(3):233–250, 1997.

N. A. Lazar. Bayesian empirical likelihood. *Biometrika*, 90:319–326, 2003.

Y. Li, T. Zeng, and J. Yu. Robust deviance information criterion for latent variable models. *Research Collection School of Economics (Open Access)*, 2012. Paper 1403.

B. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:220–239, 1988.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer series in Statistics, New York, 2002.

I. L. Macdonald and W. Zucchini. *Hidden Markov models for time series. An Introduction Using R*. CRC Press, 2009.

K. V. Mardia, G. Hughes, and C. C. Taylor. Efficiency of the pseudo-likelihood for multivariate normal and von Mises distributions. *Technical report, Stat. Dept., University of Leeds*, 2007.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavare. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.

A. Mason, S. Richardson, and N. Best. Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses. *Bayesian Analysis*, 7(1):109–146, 2012.

P. McCullagh. The conditional distribution of goodness-of-fit statistics for discrete data. *JASA*, 81(393):104–107, 1986.

C. E. McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *JASA*, 92(437):162–170, 1997.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

J. Monahan and D. Boos. Proper likelihood for Bayesian analysis. *Biometrika*, 79:271–278, 1992.

C. T. Ng, H. Joe, D. Karlis, and J. Liu. Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica*, 21:279–305, 2011.

M. Oh and Y. Lim. Bayesian analysis of time series Poisson data. *J. Applied Stat.*, 82(2): 259–271, 2001.

F. Pauli, W. Racugno, and L. Ventura. Bayesian composite marginal likelihoods. *Statistica Sinica*, 21:149–164, 2011.

M. Plummer. Discussion of the paper by Spiegelhalter et. al. *JRSS(B)*, 64:620, 2002.

J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaum, and M. W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999.

M. Ribatet, D. Cooley, and A. C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 2012.

J. F. Richard and W. Zhang. Efficient high-dimensional importance sampling. *J. Econometrics*, 141:1385–1411, 2007.

T. Ryden. Consistent and asymptotically normal parameter estimates for hidden Markov models. *The Annals of Statistics*, 22(4):1884–1895, 1994.

G. I. Sermaidis. Modeling time series of counts with an application on daily car accidents. Master thesis, Athens University of Economics and Business, 2006.

B. A. Shaby. The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, 23(3):853–876, 2014.

N. Shephard and M. K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667, 1997.

E. L. Smith and A. G. Stephenson. An extended Gaussian max-stable process model for spatial extremes. *Journal of Statistical Planning and Inference*, 139:1266–1275, 2009.

R. L. Smith. Max-stable processes and spatial extremes. Unpublished, 1990.

DJ. Spiegelhalter, NG. Best, BP. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit (with discussion). *JRSS(B)*, 64:583–639, 2002.

L. Tierney. Markov chain for exploring posterior distributions. *Ann. Stat*, 21:1701–1762, 1994.

C. Varin and P. Vidoni. Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28(1):170–185, 2009.

C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42, 2011.

G. Walker. On periodicity in series of related terms. *Proceeding of the Royal Society of London. A.*, 131:518–532, 1931.

J. Wise. Stationarity conditions for stochastic processes of the autoregressive and moving-average type. *Biometrika*, 43:215–219, 1956.

G. U. Yule. On a method of investigating periodicity in distributed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. A.*, 226:267–298, 1927.

S. L. Zeger. A regression model for time series of counts. *Biometrika*, 75:621–629, 1988.

Appendix A

Complete Results for Chapter 3

Table A.1: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(2)$ lantent process ; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.064	−0.082	−0.061	−0.080	0.000	−0.001	−0.001	−0.001	−0.002	−0.002	−0.002	−0.002
	(0.317)	(0.369)	(0.408)	(0.518)	(0.066)	(0.066)	(0.087)	(0.087)	(0.041)	(0.041)	(0.051)	(0.051)
β_1	−0.003	−0.004	−0.026	−0.020	0.000	0.000	−0.003	−0.003	0.000	0.000	−0.001	−0.001
	(0.237)	(0.247)	(0.309)	(0.318)	(0.047)	(0.047)	(0.060)	(0.060)	(0.026)	(0.026)	(0.034)	(0.034)
β_2	0.058	0.063	0.051	0.052	0.006	0.006	0.000	0.000	0.002	0.002	0.002	0.002
	(0.352)	(0.366)	(0.417)	(0.437)	(0.073)	(0.073)	(0.096)	(0.096)	(0.044)	(0.044)	(0.054)	(0.054)
ϕ_1	−0.127	−0.102	−0.054	−0.030	0.016	−0.007	0.008	0.008	0.028	0.031	0.000	−0.001
	(0.681)	(0.657)	(0.636)	(0.631)	(0.463)	(0.445)	(0.143)	(0.143)	(0.261)	(0.261)	(0.083)	(0.083)
$\boldsymbol{\phi}_2$	−0.303	−0.302	−0.269	−0.265	−0.213	−0.202	−0.022	−0.022	−0.089	−0.086	−0.001	−0.001
	(0.528)	(0.535)	(0.497)	(0.510)	(0.394)	(0.381)	(0.149)	(0.149)	(0.232)	(0.229)	(0.083)	(0.083)
σ_V	−0.067	−0.008	−0.086	−0.049	−0.052	−0.043	−0.010	−0.009	−0.015	−0.015	−0.003	−0.003
	(0.164)	(0.150)	(0.254)	(0.229)	(0.080)	(0.071)	(0.044)	(0.044)	(0.039)	(0.038)	(0.025)	(0.025)

Table A.2: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(2)$ lantent process ; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.066	−0.079	−0.020	−0.007	−0.004	−0.005	0.013	0.013	−0.002	−0.003	0.004	0.003
	(0.330)	(0.417)	(0.440)	(0.635)	(0.077)	(0.078)	(0.109)	(0.109)	(0.043)	(0.044)	(0.059)	(0.059)
β_1	−0.004	−0.007	0.010	0.016	−0.002	−0.002	0.001	0.001	−0.002	−0.002	0.000	0.000
	(0.245)	(0.258)	(0.298)	(0.308)	(0.049)	(0.049)	(0.061)	(0.061)	(0.028)	(0.028)	(0.036)	(0.036)
β_2	0.049	0.059	0.025	0.030	0.003	0.004	−0.006	−0.006	0.000	0.000	−0.003	−0.003
	(0.347)	(0.366)	(0.437)	(0.455)	(0.077)	(0.077)	(0.093)	(0.094)	(0.044)	(0.044)	(0.053)	(0.053)
ϕ_1	−0.337	−0.298	−0.152	−0.125	0.106	0.108	0.024	0.025	0.082	0.076	0.008	0.008
	(0.740)	(0.717)	(0.648)	(0.637)	(0.394)	(0.384)	(0.149)	(0.149)	(0.250)	(0.244)	(0.084)	(0.084)
ϕ_2	−0.418	−0.406	−0.368	−0.354	−0.158	−0.154	−0.033	−0.033	−0.085	−0.079	−0.010	−0.010
	(0.599)	(0.602)	(0.558)	(0.559)	(0.351)	(0.347)	(0.148)	(0.148)	(0.221)	(0.216)	(0.082)	(0.082)
σ_V	−0.058	−0.001	−0.087	−0.044	−0.027	−0.024	−0.009	−0.009	−0.011	−0.009	−0.001	−0.001
	(0.166)	(0.154)	(0.247)	(0.224)	(0.064)	(0.061)	(0.045)	(0.045)	(0.038)	(0.038)	(0.025)	(0.025)

Table A.3: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(2)$ lantent process ; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.086	−0.091	−0.041	−0.198	0.018	−0.040	−0.028	0.005	0.030	0.016	0.084	0.022
	(0.635)	(0.822)	(1.061)	(1.449)	(0.353)	(0.381)	(0.732)	(0.646)	(0.192)	(0.170)	(0.425)	(0.341)
β_1	−0.006	−0.006	−0.040	−0.030	−0.005	−0.004	−0.003	−0.003	−0.003	−0.003	−0.002	−0.002
	(0.265)	(0.278)	(0.404)	(0.411)	(0.051)	(0.051)	(0.062)	(0.062)	(0.027)	(0.027)	(0.034)	(0.034)
β_2	0.046	0.045	0.002	0.029	0.001	0.001	0.008	0.009	0.002	0.002	−0.001	−0.001
	(0.383)	(0.398)	(0.517)	(0.529)	(0.072)	(0.072)	(0.089)	(0.089)	(0.042)	(0.042)	(0.051)	(0.051)
ϕ_1	−0.560	−0.528	−0.287	−0.230	0.011	0.009	0.004	0.005	0.017	0.025	0.000	0.000
	(0.851)	(0.835)	(0.618)	(0.591)	(0.296)	(0.295)	(0.149)	(0.149)	(0.215)	(0.218)	(0.085)	(0.085)
ϕ_2	−0.273	−0.262	−0.221	−0.216	−0.018	−0.017	−0.005	−0.009	−0.019	−0.027	−0.001	−0.004
	(0.511)	(0.519)	(0.467)	(0.474)	(0.287)	(0.286)	(0.148)	(0.149)	(0.207)	(0.210)	(0.085)	(0.084)
σ_V	0.003	0.055	0.052	0.097	0.000	0.001	−0.001	−0.001	−0.004	−0.005	0.001	0.000
	(0.183)	(0.185)	(0.280)	(0.287)	(0.054)	(0.054)	(0.047)	(0.047)	(0.037)	(0.037)	(0.027)	(0.027)

Table A.4: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(3)$ lantent process ; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.20, 0.10, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.047	−0.076	−0.039	−0.074	0.006	0.005	−0.003	−0.003	0.001	0.001	−0.004	−0.004
	(0.348)	(0.474)	(0.455)	(0.629)	(0.084)	(0.084)	(0.101)	(0.102)	(0.048)	(0.048)	(0.060)	(0.060)
β_1	−0.009	−0.010	−0.018	−0.014	−0.002	−0.002	0.001	0.001	0.000	0.000	−0.001	−0.001
	(0.262)	(0.285)	(0.341)	(0.359)	(0.055)	(0.055)	(0.068)	(0.068)	(0.030)	(0.030)	(0.040)	(0.040)
β_2	0.032	0.034	0.000	0.012	−0.010	−0.010	0.006	0.006	−0.001	−0.001	0.005	0.005
	(0.374)	(0.404)	(0.476)	(0.508)	(0.088)	(0.087)	(0.101)	(0.101)	(0.048)	(0.048)	(0.059)	(0.060)
ϕ_1	−0.165	−0.135	−0.052	−0.010	0.046	0.042	0.001	0.001	0.011	0.010	0.007	0.006
	(0.711)	(0.696)	(0.645)	(0.634)	(0.220)	(0.215)	(0.122)	(0.122)	(0.120)	(0.119)	(0.069)	(0.068)
ϕ_2	−0.328	−0.321	−0.287	−0.278	−0.039	−0.033	−0.014	−0.014	−0.015	−0.013	−0.004	−0.003
	(0.600)	(0.608)	(0.561)	(0.568)	(0.234)	(0.227)	(0.126)	(0.126)	(0.130)	(0.129)	(0.073)	(0.073)
ϕ_3	−0.081	−0.070	−0.095	−0.079	−0.017	−0.020	−0.008	−0.007	−0.002	−0.002	−0.005	−0.005
	(0.403)	(0.416)	(0.396)	(0.405)	(0.200)	(0.200)	(0.118)	(0.118)	(0.113)	(0.112)	(0.067)	(0.067)
σ_V	−0.147	−0.048	−0.134	−0.070	−0.022	−0.021	−0.007	−0.007	−0.008	−0.007	−0.002	−0.002
	(0.248)	(0.196)	(0.299)	(0.260)	(0.053)	(0.052)	(0.045)	(0.045)	(0.028)	(0.027)	(0.026)	(0.026)

Table A.5: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(3)$ lantent process ; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.40, 0.20, 0.10) and (1)$\sigma_V = 0.2$; (2)$\sigma_V = 0.4$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.096	−0.104	−0.032	−0.058	−0.007	−0.008	−0.002	−0.004	−0.002	−0.003	0.002	0.002
	(0.409)	(0.615)	(0.551)	(0.778)	(0.109)	(0.109)	(0.147)	(0.148)	(0.059)	(0.059)	(0.083)	(0.083)
β_1	0.012	0.012	−0.065	−0.064	−0.005	−0.005	0.002	0.002	0.001	0.001	−0.002	−0.002
	(0.259)	(0.278)	(0.360)	(0.376)	(0.057)	(0.057)	(0.071)	(0.071)	(0.033)	(0.033)	(0.040)	(0.040)
β_2	0.068	0.072	−0.004	0.004	−0.001	−0.001	−0.008	−0.008	0.000	−0.001	−0.005	−0.005
	(0.385)	(0.415)	(0.482)	(0.510)	(0.084)	(0.084)	(0.101)	(0.101)	(0.047)	(0.047)	(0.059)	(0.059)
ϕ_1	−0.252	−0.196	−0.138	−0.106	0.077	0.075	0.006	0.007	0.043	0.043	0.004	0.004
	(0.746)	(0.720)	(0.645)	(0.632)	(0.244)	(0.238)	(0.121)	(0.120)	(0.132)	(0.132)	(0.072)	(0.072)
ϕ_2	−0.408	−0.394	−0.315	−0.305	−0.088	−0.085	−0.020	−0.020	−0.050	−0.050	−0.001	−0.001
	(0.656)	(0.657)	(0.588)	(0.593)	(0.289)	(0.284)	(0.145)	(0.144)	(0.166)	(0.165)	(0.085)	(0.085)
ϕ_3	−0.084	−0.066	−0.064	−0.049	0.010	0.009	0.006	0.006	0.008	0.009	−0.004	−0.004
	(0.406)	(0.413)	(0.387)	(0.397)	(0.191)	(0.190)	(0.119)	(0.118)	(0.113)	(0.113)	(0.069)	(0.069)
σ_V	−0.136	−0.045	−0.109	−0.049	−0.020	−0.019	−0.006	−0.005	−0.007	−0.007	−0.003	−0.003
	(0.242)	(0.199)	(0.284)	(0.252)	(0.056)	(0.055)	(0.046)	(0.046)	(0.031)	(0.031)	(0.026)	(0.027)

Table A.6: Bias and RMSE (in parenthesis) of parameter estimates for SM and PGS methods with $AR(3)$ lantent process ; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.88, 0.34, −0.28) and (1)$\sigma_V = 0.2$; (2)$\sigma_V = 0.4$.

	n=20				n=365				n=1095			
	(1)		(2)		(1)		(2)		(1)		(2)	
Par.	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS	SM	PGS
β_0	−0.122	−0.079	−0.059	−0.134	0.047	−0.024	0.209	0.137	0.093	−0.001	0.105	0.103
	(0.927)	(1.314)	(1.425)	(1.801)	(0.576)	(0.440)	(0.875)	(0.653)	(0.296)	(0.217)	(0.441)	(0.378)
β_1	−0.020	−0.021	−0.004	−0.009	0.003	0.003	−0.004	−0.004	0.000	0.000	0.000	0.001
	(0.333)	(0.351)	(0.426)	(0.453)	(0.056)	(0.056)	(0.067)	(0.067)	(0.032)	(0.032)	(0.039)	(0.038)
β_2	0.026	0.031	0.068	0.091	0.002	0.002	0.001	−0.001	0.003	0.003	0.000	−0.001
	(0.469)	(0.494)	(0.582)	(0.629)	(0.081)	(0.080)	(0.091)	(0.090)	(0.044)	(0.044)	(0.052)	(0.052)
ϕ_1	−0.272	−0.228	−0.198	−0.163	0.080	0.082	0.012	0.011	0.030	0.030	−0.003	−0.003
	(0.682)	(0.653)	(0.593)	(0.573)	(0.215)	(0.215)	(0.109)	(0.108)	(0.105)	(0.104)	(0.061)	(0.060)
ϕ_2	−0.443	−0.429	−0.408	−0.399	−0.157	−0.155	−0.035	−0.033	−0.056	−0.052	−0.001	0.000
	(0.716)	(0.716)	(0.670)	(0.673)	(0.350)	(0.351)	(0.171)	(0.171)	(0.172)	(0.171)	(0.095)	(0.095)
ϕ_3	0.236	0.242	0.224	0.220	0.080	0.073	0.024	0.018	0.026	0.022	0.005	0.004
	(0.452)	(0.463)	(0.438)	(0.442)	(0.193)	(0.193)	(0.116)	(0.116)	(0.103)	(0.102)	(0.065)	(0.064)
σ_V	0.012	0.081	−0.007	0.071	−0.009	−0.010	0.003	0.002	−0.004	−0.005	0.001	0.001
	(0.244)	(0.247)	(0.301)	(0.315)	(0.055)	(0.055)	(0.049)	(0.049)	(0.029)	(0.029)	(0.027)	(0.027)

Appendix B

Complete Results for Chapter 4

Table B.1: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.002	−.004	−.003	−.004	−.004	−.002	−.001	−.002	−.002	−.001
	(.041)	(.027)	(.040)	(.045)	(.045)	(.051)	(.034)	(.058)	(.071)	(.072)
β_1	.000	.000	.000	.000	.000	−.001	−.001	−.001	−.001	−.001
	(.026)	(.017)	(.026)	(.027)	(.027)	(.034)	(.022)	(.040)	(.034)	(.034)
β_2	.002	.001	.002	.001	.001	.002	.002	.002	.002	.002
	(.044)	(.029)	(.043)	(.044)	(.044)	(.054)	(.034)	(.064)	(.054)	(.054)
ϕ_1	.028	−.007	−.060	−.008	−.008	.000	−.008	−.009	−.009	−.009
	(.261)	(.229)	(.452)	(.227)	(.234)	(.083)	(.085)	(.260)	(.085)	(.085)
ϕ_2	−.089	−.081	−.198	−.082	−.081	−.001	−.014	−.095	−.014	−.014
	(.232)	(.207)	(.377)	(.205)	(.209)	(.083)	(.083)	(.233)	(.084)	(.084)
σ_V	−.015	−.005	−.023	−.002	−.004	−.003	−.005	−.018	−.004	−.005
	(.039)	(.025)	(.045)	(.035)	(.036)	(.025)	(.017)	(.040)	(.028)	(.027)

Table B.2: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.002	−.003	−.003	−.004	−.003	.004	.005	.003	.003	.005
	(.043)	(.028)	(.044)	(.049)	(.049)	(.059)	(.037)	(.067)	(.082)	(.083)
β_1	−.002	−.002	−.002	−.002	−.002	.000	−.001	−.001	−.001	−.001
	(.028)	(.019)	(.029)	(.029)	(.029)	(.036)	(.025)	(.046)	(.039)	(.039)
β_2	.000	.000	.000	−.001	.000	−.003	−.004	−.003	−.003	−.004
	(.044)	(.029)	(.047)	(.044)	(.044)	(.053)	(.036)	(.070)	(.054)	(.054)
ϕ_1	.082	.045	−.007	.025	.040	.008	.010	.043	.008	.008
	(.250)	(.212)	(.397)	(.206)	(.224)	(.084)	(.086)	(.251)	(.087)	(.086)
ϕ_2	−.085	−.070	−.188	−.064	−.069	−.010	−.019	−.088	−.020	−.019
	(.221)	(.197)	(.370)	(.190)	(.200)	(.082)	(.087)	(.245)	(.089)	(.088)
σ_V	−.011	−.003	−.006	.002	.000	−.001	−.004	−.010	−.003	−.004
	(.038)	(.027)	(.044)	(.037)	(.039)	(.025)	(.018)	(.051)	(.029)	(.029)

Table B.3: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.030	.019	.014	.016	.019	.084	.011	.005	.011	.011
	(.192)	(.101)	(.143)	(.172)	(.173)	(.425)	(.201)	(.275)	(.342)	(.341)
β_1	−.003	−.004	−.004	−.004	−.004	−.002	−.001	−.002	−.001	−.001
	(.027)	(.022)	(.064)	(.034)	(.034)	(.034)	(.033)	(.100)	(.047)	(.047)
β_2	.002	.001	.002	.000	.001	−.001	−.001	.000	−.001	−.001
	(.042)	(.029)	(.092)	(.044)	(.043)	(.051)	(.036)	(.135)	(.054)	(.054)
ϕ_1	.017	.015	−.060	−.004	.010	.000	−.004	−.027	−.006	−.005
	(.215)	(.235)	(.410)	(.240)	(.238)	(.085)	(.093)	(.348)	(.097)	(.097)
ϕ_2	−.019	−.018	.041	−.005	−.017	−.001	−.001	.015	−.002	−.001
	(.207)	(.228)	(.403)	(.231)	(.231)	(.085)	(.092)	(.345)	(.096)	(.096)
σ_V	−.004	−.007	.017	−.004	−.006	.001	−.004	.014	−.003	−.004
	(.037)	(.033)	(.072)	(.044)	(.042)	(.027)	(.024)	(.096)	(.031)	(.031)

Table B.4: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.000	−.002	−.002	−.003	−.002	−.001	.002	.000	.000	.002
	(.066)	(.042)	(.061)	(.069)	(.068)	(.087)	(.056)	(.086)	(.099)	(.099)
β_1	.000	.000	.000	.000	.000	−.003	−.002	−.002	−.002	−.002
	(.047)	(.032)	(.044)	(.047)	(.047)	(.060)	(.040)	(.060)	(.060)	(.060)
β_2	.006	.005	.006	.005	.005	.000	−.002	−.001	−.001	−.002
	(.073)	(.046)	(.068)	(.074)	(.073)	(.096)	(.064)	(.097)	(.096)	(.096)
ϕ_1	.016	−.037	−.127	−.041	−.043	.008	−.003	−.029	−.006	−.005
	(.463)	(.400)	(.570)	(.367)	(.412)	(.143)	(.145)	(.351)	(.146)	(.145)
ϕ_2	−.213	−.211	−.273	−.189	−.199	−.022	−.046	−.148	−.047	−.046
	(.394)	(.356)	(.468)	(.333)	(.363)	(.149)	(.148)	(.305)	(.150)	(.150)
σ_V	−.052	−.028	−.046	−.006	−.007	−.010	−.015	−.032	−.011	−.013
	(.080)	(.046)	(.070)	(.063)	(.103)	(.044)	(.031)	(.062)	(.046)	(.046)

Table B.5: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.004	−.005	−.005	−.006	−.005	.013	.013	.011	.011	.013
	(.077)	(.051)	(.072)	(.079)	(.079)	(.109)	(.072)	(.108)	(.121)	(.122)
β_1	−.002	−.002	−.002	−.002	−.002	.001	.001	.001	.001	.001
	(.049)	(.034)	(.049)	(.051)	(.051)	(.061)	(.043)	(.070)	(.067)	(.067)
β_2	.003	.003	.004	.003	.003	−.006	−.006	−.005	−.006	−.006
	(.077)	(.051)	(.076)	(.078)	(.077)	(.093)	(.065)	(.109)	(.096)	(.096)
ϕ_1	.106	.062	−.155	.017	.040	.024	.025	.026	.018	.021
	(.394)	(.343)	(.556)	(.326)	(.377)	(.149)	(.149)	(.335)	(.149)	(.148)
ϕ_2	−.158	−.141	−.307	−.127	−.130	−.033	−.047	−.151	−.050	−.047
	(.351)	(.314)	(.485)	(.297)	(.331)	(.148)	(.149)	(.322)	(.153)	(.153)
σ_V	−.027	−.016	−.019	.010	.003	−.009	−.014	−.016	−.010	−.013
	(.064)	(.046)	(.061)	(.080)	(.095)	(.045)	(.034)	(.071)	(.049)	(.049)

Table B.6: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.85, 0.10)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.018	−.002	−.014	−.009	−.002	−.028	.007	−.010	.012	.007
	(.353)	(.163)	(.237)	(.275)	(.277)	(.732)	(.307)	(.439)	(.534)	(.537)
β_1	−.005	−.005	−.005	−.004	−.005	−.003	−.004	−.005	−.003	−.004
	(.051)	(.042)	(.112)	(.064)	(.063)	(.062)	(.058)	(.177)	(.083)	(.083)
β_2	.001	.000	.004	−.001	.000	.008	.007	.012	.007	.007
	(.072)	(.049)	(.157)	(.074)	(.073)	(.089)	(.066)	(.239)	(.095)	(.095)
ϕ_1	.011	−.002	−.124	−.053	−.033	.004	.000	−.057	−.008	−.004
	(.296)	(.314)	(.438)	(.321)	(.344)	(.149)	(.1640	(.399)	(.169)	(.167)
ϕ_2	−.018	−.011	.038	.021	−.006	−.005	−.009	.021	−.011	−.009
	(.287)	(.305)	(.405)	(.307)	(.326)	(.148)	(.162)	(.390)	(.166)	(.165)
σ_V	.000	−.007	.069	.008	.001	−.001	−.011	.066	−.008	−.010
	(.054)	(.049)	(.125)	(.069)	(.070)	(.047)	(.043)	(.160)	(.053)	(.051)

Table B.7: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.19, 0.0.06)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.064	−.058	−.069	−.064	−.058	−.061	−.060	−.071	−.067	−.060
	(.317)	(.270)	(.302)	(.321)	(.317)	(.408)	(.325)	(.365)	(.388)	(.384)
β_1	−.003	−.003	−.002	−.003	−.003	−.026	−.019	−.019	−.018	−.019
	(.237)	(.197)	(.219)	(.238)	(.235)	(.309)	(.254)	(.283)	(.302)	(.299)
β_2	.058	.054	.059	.055	.054	.051	.053	.058	.055	.053
	(.352)	(.303)	(.338)	(.360)	(.356)	(.417)	(.347)	(.390)	(.419)	(.416)
ϕ_1	−.127	−.159	−.160	−.168	−.154	−.054	−.123	−.121	−.133	−.120
	(.681)	(.658)	(.668)	(.637)	(.646)	(.636)	(.615)	(.621)	(.600)	(.584)
ϕ_2	−.303	−.309	−.310	−.291	−.370	−.269	−.307	−.310	−.290	−.370
	(.528)	(.530)	(.533)	(.567)	(.384)	(.497)	(.527)	(.531)	(.566)	(.377)
σ_V	−.067	−.064	−.045	.040	−.036	−.086	−.126	−.101	−.006	−.105
	(.164)	(.129)	(.133)	(.199)	(.159)	(.254)	(.204)	(.212)	(.259)	(.219)

Table B.8: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20)$ and $(1)\sigma_V = 0.2$; $(2)\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.066	−.065	−.076	−.071	−.065	−.020	−.006	−.019	−.014	−.006
	(.330)	(.282)	(.314)	(.333)	(.329)	(.440)	(.357)	(.398)	(.420)	(.414)
β_1	−.004	−.005	−.004	−.004	−.005	.010	.014	.014	.014	.014
	(.245)	(.205)	(.228)	(.246)	(.242)	(.298)	(.248)	(.279)	(.296)	(.293)
β_2	.049	.050	.055	.050	.050	.025	.020	.026	.023	.020
	(.347)	(.296)	(.331)	(.355)	(.352)	(.437)	(.364)	(.411)	(.437)	(.430)
ϕ_1	−.337	−.368	−.365	−.366	−.361	−.152	−.235	−.248	−.231	−.226
	(.740)	(.732)	(.741)	(.717)	(.717)	(.648)	(.636)	(.651)	(.623)	(.597)
ϕ_2	−.418	−.451	−.449	−.432	−.512	−.368	−.449	−.449	−.429	−.512
	(.599)	(.620)	(.624)	(.649)	(.521)	(.558)	(.619)	(.624)	(.649)	(.516)
σ_V	−.058	−.060	−.040	.047	−.029	−.087	−.126	−.101	−.005	−.107
	(.166)	(.129)	(.135)	(.205)	(.165)	(.247)	(.200)	(.209)	(.254)	(.215)

Table B.9: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(2)$ lantent process and a sample size of $n = 20$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.85, 0.10) and (1)$\sigma_V = 0.2$; (2)$\sigma_V = 0.4$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.086	−.074	−.090	−.081	−.074	−.041	−.155	−.212	−.204	−.155
	(.635)	(.586)	(.620)	(.630)	(.627)	(1.061)	(1.001)	(1.116)	(1.124)	(1.075)
β_1	−.006	−.009	−.010	−.009	−.009	−.040	−.044	−.042	−.042	−.044
	(.265)	(.222)	(.255)	(.268)	(.265)	(.404)	(.370)	(.466)	(.441)	(.423)
β_2	.046	.044	.051	.046	.044	.002	.004	.012	.013	.004
	(.383)	(.329)	(.380)	(.398)	(.393)	(.517)	(.446)	(.591)	(.557)	(.537)
ϕ_1	−.560	−.602	−.641	−.611	−.594	−.287	−.334	−.425	−.337	−.314
	(.851)	(.852)	(.889)	(.854)	(.820)	(.618)	(.625)	(.685)	(.639)	(.559)
ϕ_2	−.273	−.346	−.353	−.331	−.408	−.221	−.347	−.351	−.333	−.410
	(.511)	(.551)	(.559)	(.587)	(.418)	(.467)	(.550)	(.557)	(.589)	(.413)
σ_V	.003	−.011	.018	.100	.019	.052	.000	.103	.226	.023
	(.183)	(.137)	(.161)	(.237)	(.187)	(.280)	(.222)	(.322)	(.463)	(.267)

Table B.10: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.20, 0.10, 0.10) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.001	.001	.001	.001	.001	−.004	−.003	−.005	−.005	−.003
	(.048)	(.032)	(.050)	(.061)	(.061)	(.060)	(.040)	(.068)	(.089)	(.089)
β_1	.000	.001	.001	.001	.001	−.001	−.002	−.002	−.002	−.002
	(.030)	(.019)	(.032)	(.030)	(.030)	(.040)	(.027)	(.047)	(.041)	(.041)
β_2	−.001	−.001	−.001	−.001	−.001	.005	.005	.006	.006	.005
	(.048)	(.032)	(.053)	(.049)	(.049)	(.059)	(.040)	(.073)	(.061)	(.061)
ϕ_1	.011	.005	.032	.000	.002	.007	.004	.032	.002	.002
	(.120)	(.121)	(.326)	(.119)	(.120)	(.069)	(.071)	(.233)	(.070)	(.069)
ϕ_2	−.015	−.025	−.131	−.025	−.025	−.004	−.012	−.082	−.013	−.012
	(.130)	(.132)	(.328)	(.126)	(.126)	(.073)	(.079)	(.246)	(.073)	(.073)
ϕ_3	−.002	−.009	−.012	−.008	−.009	−.005	−.011	−.007	−.011	−.011
	(.113)	(.114)	(.242)	(.112)	(.113)	(.067)	(.072)	(.192)	(.071)	(.071)
σ_V	−.008	−.007	−.027	−.005	−.006	−.002	−.007	−.024	−.005	−.006
	(.028)	(.020)	(.047)	(.029)	(.029)	(.026)	(.019)	(.048)	(.033)	(.033)

Table B.11: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.002	−.002	−.003	−.003	−.002	.002	.004	.002	.001	.004
	(.059)	(.037)	(.059)	(.074)	(.074)	(.083)	(.050)	(.085)	(.110)	(.110)
β_1	.001	.001	.001	.001	.001	−.002	−.004	−.004	−.004	−.004
	(.033)	(.025)	(.040)	(.037)	(.037)	(.040)	(.031)	(.057)	(.047)	(.047)
β_2	.000	−.001	.000	−.001	−.001	−.005	−.005	−.005	−.005	−.005
	(.047)	(.031)	(.058)	(.048)	(.048)	(.059)	(.041)	(.083)	(.061)	(.061)
ϕ_1	.043	.044	.126	.035	.039	.004	.010	.089	.005	.006
	(.132)	(.142)	(.374)	(.130)	(.131)	(.072)	(.080)	(.294)	(.074)	(.074)
ϕ_2	−.050	−.064	−.183	−.061	−.062	−.001	−.013	−.139	−.011	−.011
	(.166)	(.188)	(.450)	(.162)	(.164)	(.085)	(.107)	(.373)	(.085)	(.086)
ϕ_3	.008	.015	.053	.015	.015	−.004	−.002	.049	−.003	−.002
	(.113)	(.124)	(.276)	(.114)	(.115)	(.069)	(.080)	(.232)	(.072)	(.072)
σ_V	−.007	−.009	−.030	−.007	−.008	−.003	−.010	−.032	−.007	−.009
	(.031)	(.025)	(.061)	(.035)	(.034)	(.026)	(.022)	(.067)	(.037)	(.037)

Table B.12: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 1095$; parameters $\boldsymbol{\beta}$=(1.5, −0.1, 0.4), $\boldsymbol{\phi}$=(0.88, 0.34, −0.28) and (1)$\sigma_V = 0.3$; (2)$\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.093	.006	.003	.009	.006	.105	−.018	−.023	−.002	−.018
	(.296)	(.132)	(.181)	(.232)	(.234)	(.441)	(.212)	(.292)	(.375)	(.381)
β_1	.000	.000	−.001	.000	.000	.000	.001	.001	.002	.001
	(.032)	(.029)	(.073)	(.043)	(.043)	(.039)	(.035)	(.106)	(.054)	(.054)
β_2	.003	.003	.003	.000	.003	.000	−.003	−.001	−.004	−.003
	(.044)	(.032)	(.098)	(.050)	(.049)	(.052)	(.040)	(.143)	(.061)	(.061)
ϕ_1	.030	.060	.115	.042	.043	−.003	.013	.118	.001	.004
	(.105)	(.146)	(.417)	(.117)	(.121)	(.061)	(.077)	(.378)	(.067)	(.071)
ϕ_2	−.056	−.113	−.381	−.092	−.097	−.001	−.036	−.362	−.025	−.028
	(.172)	(.277)	(.695)	(.205)	(.196)	(.095)	(.165)	(.664)	(.113)	(.111)
ϕ_3	.026	.051	.267	.044	.049	.005	.021	.243	.018	.020
	(.103)	(.162)	(.434)	(.125)	(.127)	(.065)	(.111)	(.408)	(.087)	(.093)
σ_V	−.004	−.013	−.002	−.015	−.012	.001	−.007	−.001	−.009	−.007
	(.029)	(.032)	(.093)	(.053)	(.039)	(.027)	(.029)	(.126)	(.042)	(.035)

Table B.13: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.20, 0.10, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.006	.006	.005	.005	.006	−.003	.000	−.002	−.002	.000
	(.084)	(.056)	(.078)	(.091)	(.091)	(.101)	(.063)	(.098)	(.119)	(.119)
β_1	−.002	−.003	−.003	−.003	−.003	.001	.001	.000	.000	.001
	(.055)	(.037)	(.053)	(.055)	(.055)	(.068)	(.047)	(.070)	(.070)	(.070)
β_2	−.010	−.011	−.010	−.011	−.011	.006	.005	.005	.005	.005
	(.088)	(.060)	(.086)	(.088)	(.088)	(.101)	(.065)	(.106)	(.102)	(.102)
ϕ_1	.046	.023	.001	.013	.017	.001	−.002	.015	−.007	−.006
	(.220)	(.207)	(.444)	(.201)	(.206)	(.122)	(.129)	(.319)	(.126)	(.126)
ϕ_2	−.039	−.050	−.187	−.050	−.049	−.014	−.031	−.131	−.031	−.031
	(.234)	(.220)	(.419)	(.209)	(.212)	(.126)	(.135)	(.320)	(.125)	(.126)
ϕ_3	−.017	−.025	−.053	−.027	−.025	−.008	−.014	−.021	−.014	−.013
	(.200)	(.192)	(.310)	(.189)	(.192)	(.118)	(.123)	(.242)	(.122)	(.122)
σ_V	−.022	−.019	−.044	−.013	−.016	−.007	−.017	−.041	−.014	−.016
	(.053)	(.037)	(.071)	(.053)	(.054)	(.045)	(.035)	(.071)	(.052)	(.052)

Table B.14: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.007	−.007	−.009	−.009	−.007	−.002	−.002	−.005	−.005	−.002
	(.109)	(.071)	(.099)	(.114)	(.115)	(.147)	(.091)	(.135)	(.159)	(.160)
β_1	−.005	−.005	−.005	−.005	−.005	.002	.003	.003	.003	.003
	(.057)	(.042)	(.064)	(.063)	(.063)	(.071)	(.056)	(.089)	(.083)	(.083)
β_2	−.001	−.002	−.001	−.002	−.002	−.008	−.008	−.007	−.008	−.008
	(.084))	(.057)	(.094)	(.085)	(.085)	(.101)	(.070)	(.125)	(.103)	(.103)
ϕ_1	.077	.065	.091	.040	.050	.006	.020	.090	.005	.009
	(.244)	(.237)	(.446)	(.213)	(.221)	(.121)	(.137)	(.366)	(.123)	(.122)
ϕ_2	−.088	−.103	−.220	−.094	−.096	−.020	−.050	−.181	−.045	−.045
	(.289)	(.298)	(.507)	(.256)	(.262)	(.145)	(.180)	(.444)	(.143)	(.142)
ϕ_3	.010	.022	.040	.020	.022	.006	.012	.052	.010	.013
	(.191)	(.196)	(.320)	(.182)	(.185)	(.119)	(.132)	(.280)	(.119)	(.120)
σ_V	−.020	−.022	−.036	−.014	−.018	−.006	−.018	−.040	−.014	−.017
	(.056)	(.044)	(.079)	(.061)	(.062)	(.046)	(.038)	(.093)	(.059)	(.057)

Table B.15: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 365$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.88, 0.34, -0.28)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	.047	−.022	−.031	−.031	−.022	.209	.003	−.011	.022	.003
	(.576)	(.219)	(.299)	(.374)	(.375)	(.875)	(.331)	(.455)	(.573)	(.578)
β_1	.003	.007	.007	.008	.007	−.004	.000	−.001	.001	.000
	(.056)	(.054)	(.126)	(.080)	(.079)	(.067)	(.070)	(.179)	(.101)	(.101)
β_2	.002	.002	.004	−.004	.002	.001	−.001	.002	−.005	−.001
	(.081)	(.060)	(.167)	(.090)	(.087)	(.091)	(.069)	(.232)	(.104)	(.103)
ϕ_1	.080	.115	.007	.077	.069	.012	.049	.045	.019	.020
	(.215)	(.253)	(.459)	(.206)	(.230)	(.109)	(.155)	(.428)	(.121)	(.145)
ϕ_2	−.157	−.224	−.371	−.193	−.192	−.035	−.121	−.372	−.093	−.098
	(.350)	(.456)	(.695)	(.359)	(.325)	(.171)	(.307)	(.685)	(.213)	(.195)
ϕ_3	.080	.104	.349	.089	.098	.024	.065	.313	.054	.061
	(.193)	(.256)	(.508)	(.214)	(.200)	(.116)	(.192)	(.474)	(.152)	(.153)
σ_V	−.009	−.024	.038	−.029	−.018	.003	−.017	.045	−.024	−.015
	(.055)	(.056)	(.125)	(.103)	(.071)	(.049)	(.055)	(.173)	(.094)	(.065)

Table B.16: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.20, 0.10, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.047	−.045	−.052	−.056	−.036	−.039	−.056	−.064	−.065	−.061
	(.348)	(.301)	(.321)	(.359)	(.390)	(.455)	(.389)	(.413)	(.453)	(.497)
β_1	−.009	−.012	−.012	−.010	−.028	−.018	−.019	−.019	−.016	−.021
	(.262)	(.225)	(.238)	(.269)	(.280)	(.341)	(.283)	(.303)	(.338)	(.347)
β_2	.032	.025	.029	.031	.027	.000	.013	.016	.013	.034
	(.374)	(.323)	(.345)	(.389)	(.426)	(.476)	(.411)	(.439)	(.487)	(.523)
ϕ_1	−.165	−.220	−.216	−.209	−.279	−.052	−.115	−.114	−.113	−.178
	(.711)	(.682)	(.689)	(.674)	(.747)	(.645)	(.617)	(.622)	(.614)	(.600)
ϕ_2	−.328	−.355	−.351	−.328	−.333	−.287	−.348	−.348	−.337	−.281
	(.600)	(.612)	(.614)	(.645)	(.674)	(.561)	(.603)	(.605)	(.644)	(.529)
ϕ_3	−.081	−.106	−.102	−.101	−.101	−.095	−.105	−.098	−.096	−.137
	(.403)	(.419)	(.420)	(.459)	(.447)	(.396)	(.418)	(.419)	(.456)	(.416)
σ_V	−.147	−.117	−.108	.035	.039	−.134	−.178	−.164	.009	−.096
	(.248)	(.174)	(.177)	(.240)	(.382)	(.299)	(.241)	(.243)	(.295)	(.297)

Table B.17: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.40, 0.20, 0.10)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.096	−.100	−.107	−.106	−.046	−.032	−.028	−.037	−.046	.002
	(.409)	(.364)	(.385)	(.419)	(.460)	(.551)	(.455)	(.483)	(.527)	(.560)
β_1	.012	.009	.009	.011	−.009	−.065	−.060	−.060	−.059	−.056
	(.259)	(.221)	(.237)	(.268)	(.285)	(.360)	(.312)	(.333)	(.362)	(.371)
β_2	.068	.067	.071	.070	.068	−.004	−.010	−.007	.003	.007
	(.385)	(.341)	(.366)	(.407)	(.443)	(.482)	(.407)	(.442)	(.491)	(.517)
ϕ_1	−.252	−.286	−.300	−.285	−.368	−.138	−.191	−.209	−.187	−.303
	(.746)	(.717)	(.726)	(.716)	(.801)	(.645)	(.635)	(.639)	(.633)	(.683)
ϕ_2	−.408	−.452	−.449	−.428	−.387	−.315	−.453	−.443	−.429	−.336
	(.656)	(.675)	(.676)	(.706)	(.712)	(.588)	(.671)	(.668)	(.702)	(.588)
ϕ_3	−.084	−.107	−.104	−.099	−.114	−.064	−.106	−.101	−.102	−.106
	(.406)	(.418)	(.422)	(.458)	(.449)	(.387)	(.420)	(.420)	(.458)	(.405)
σ_V	−.136	−.122	−.110	.029	.027	−.109	−.167	−.148	.034	−.088
	(.242)	(.180)	(.182)	(.242)	(.368)	(.284)	(.235)	(.237)	(.302)	(.311)

Table B.18: Bias and RMSE (in parenthesis) of parameter estimates for Bayesian methods using Full and composite likelihood with $AR(3)$ latent process and sample size of $n = 20$; parameters $\boldsymbol{\beta}$=$(1.5, -0.1, 0.4)$, $\boldsymbol{\phi}$=$(0.88, 0.34, -0.28)$ and $(1)\sigma_V = 0.3$; $(2)\sigma_V = 0.5$.

	Parameter Set (1)					Parameter Set (2)				
Par.	Full	Comp	Mag.	Curv.	OFS	Full	Comp	Mag.	Curv.	OFS
β_0	−.122	−.104	−.122	−.119	.061	−.059	−.178	−.260	−.195	.131
	(.927)	(.803)	(.855)	(.891)	(.863)	(1.425)	(1.270)	(1.417)	(1.399)	(1.317)
β_1	−.020	−.028	−.031	−.023	−.030	−.004	−.043	−.064	−.039	−.040
	(.333)	(.311)	(.361)	(.365)	(.413)	(.426)	(.430)	(.567)	(.487)	(.572)
β_2	.026	.022	.029	.040	.053	.068	.078	.116	.084	.091
	(.469)	(.410)	(.486)	(.493)	(.467)	(.582)	(.530)	(.713)	(.641)	(.621)
ϕ_1	−.272	−.338	−.410	−.341	−.483	−.198	−.264	−.348	−.257	−.414
	(.682)	(.682)	(.713)	(.705)	(.779)	(.593)	(.627)	(.665)	(.657)	(.664)
ϕ_2	−.443	−.584	−.581	−.573	−.401	−.408	−.582	−.595	−.570	−.368
	(.716)	(.793)	(.786)	(.835)	(.640)	(.670)	(.794)	(.800)	(.838)	(.542)
ϕ_3	.236	.274	.275	.284	.251	.224	.273	.279	.285	.242
	(.452)	(.488)	(.491)	(.529)	(.461)	(.438)	(.488)	(.493)	(.529)	(.428)
σ_V	.012	−.010	.040	.189	.081	−.007	−.054	.062	.230	.025
	(.244)	(.181)	(.220)	(.347)	(.333)	(.301)	(.252)	(.357)	(.476)	(.319)

Appendix C

C Codes for The SM Method

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>
//#define MATHLIB_STANDALONE
#include <S.h>
#include <Rdefines.h>
#ifndef USING_R
//#extern void F77_NAME(dqrdca)();
#else
#include <R.h>
#include <Rmath.h>
#include <R_ext/Applic.h>
#include <R_ext/Boolean.h>
#include <R_ext/RStartup.h>
#include <R_ext/Linpack.h>
#include <R_ext/Lapack.h>
#endif /* USING_R */

// The Gibbs sampler.

// col4: the y data.
// betas: a vector of length (gibiter)*(num of beta): the first gibiter elements are
             for beta_0, the next are for beta_1 and so on.

//arpar:  a vector of length (gibiter)*(p + 1): the first gibiter elements are for rstar_1 the next are for
              rstar_2 and so on. The last gibiter element are for sigma^2_V.

// gibiter: number of iterations of the gibbs sampler.
// accpetbet: a vector of length "betnum" contains the acceptance probabilites for the M-H when sampling betas.
// cols: the X data as a vector.
// par: initial values.
// accepteta: No longer used.
// ss: sample size.
// betnumber: number of betas in the model.
// acceptrst: acceptance probability for the M-H when sampling rstar.

void simulation(double *col4,double *betas,double *arpar,int *gibiter, double *acceptbet,double *cols, double *par,
                        double *accepeta, int *ss, int *betnumber, double *acceptrst)
{
    int i,j,k,l,n=*ss,betnum=*betnumber,gibit=*gibiter,p=2,dim=p*p;
    double etas[n+p],ydat,nn;
    double col[n][betnum];
    double muet,sigetsq,siget,betxt,max,fmax,etprop,u,deno;
    double sigetprop,sigprop,mubet=0,sigbetsq=10000;
    double aa,aa0, aa2, aa3, bb1;
    double aa1,bb0,bb00,bb2,bb22,bb3;
    double sumetsq,sumetsq1,sumetsq2,sumetet1,sumetet2,sumet1et2,loglik1,loglik2;
    double gam1, alfa1, cond,fb1,fb2;
    double mu1,sig1,mu2,sig2,ssum1,ssum2;
    double ratio, ratio1,ratio2, unii;

    double betprop,betasmh[betnum],sigvsqq;
    double a,mean,expbetxt,count,maxit=20000,times=0.0;
    double mode,stdev,covmat[dim],gammas[p],mvndat[p],logdens[2],logpr[1],rstar[p],rstarprop[p],phis[p+1];
    double A1,B1,a1,a2,sigma2sq,sigma3sq,rho1;
    alfa1=(double)(n-1.0)/2.0;

for(j=n;j<(n+p);j++) etas[j]=0.0;
```

```
nn=(double) n;
acceptbet[0]=0.0;
acceptrst[0]=0.0;
sigprop=0.15;

// putting the data in a matrix
for(k=0;k<betnum;k++)
{
    for(j=0;j<n;j++)
        col[j][k]=cols[k*n+j];
}

// initialize eta
GetRNGstate();
for(j=0; j<n; j++)
{
    etas[j]=0.01;
    accepeta[j]=0.0;
}
// initialize the parameters
for(j=0;j<betnum;j++) betas[j*gibit]=par[j];

for(j=0;j<=p;j++) arpar[j*gibit]=par[betnum+j];
for(j=0;j<p;j++) rstar[j]=arpar[j*gibit];
sigvsqq=arpar[p*gibit];


for(i=1;i<gibit;i++)

{
    //Sampling etas using the second method discussed in Chapter 2.

    // finding \phi and \gamma from rstar
    getphi(&phis, &rstar, &p);
    getgammas(&gammas,&phis,&sigvsqq);
    phis[2]=0.0;

    //##################
    //## SAMPLING ETA_1

    // the mean and variance for the posterior distribution of \eta_1: muet and sigsqet respectively
    deno=gammas[0]*gammas[0]-gammas[1]*gammas[1];
    a2=phis[1];
    a1=phis[0];
    rho1=gammas[1]/gammas[0];
    sigma2sq=gammas[0]*(1-rho1*rho1);
    sigma3sq=sigvsqq;

    A1=1.0/gammas[0]+rho1*rho1/sigma2sq+a2*a2/sigma3sq+phis[2]*phis[2]/sigvsqq;
    B1=rho1*etas[1]/sigma2sq+(a2*etas[2]-a1*a2*etas[1])/sigma3sq+(phis[2]*etas[3]-phis[0]*phis[2]*etas[2]-
           phis[1]*phis[2]*etas[1])/sigvsqq;
    muet = B1/A1;
    sigetsq=1.0/A1;
    siget=sqrt(sigetsq);
    sigetprop=siget;
    betxt=0.0;

    // calculationg X^T*Beta

    for(j=0;j<betnum;j++) betxt+=betas[j*gibit+i-1]*col[0][j];

    // calling y_1
    ydat=col4[0];

    // finding the mean of the envelope function: called mean
    nlm(&a,&betxt,&ydat,&muet,&sigetsq);
```

```
expbetxt=exp(betxt+a);

mean=muet+sigetsq*(ydat-expbetxt);

   //## THE ACCEPT-REJECT METHOD
count=0.0;
   do{
        etprop=rnorm(mean,siget);
        u=log(runif(0,1));
        ratio=ydat*etprop-exp(betxt+etprop)+dnorm(etprop,muet,siget,1)-((a-1)*expbetxt+(mean*mean
                -muet*muet)/(2*sigetsq)+dnorm(etprop,mean,siget,1));
       count+=1.0;
     }while(u>ratio);

  etas[0] = etprop;

    //##################
    //## SAMPLING ETA_2
// mean and variance of the posterior of \eta_2

A1=1.0/sigma2sq+a1*a1/sigma3sq+(phis[1]*phis[1]+phis[2]*phis[2])/sigvsqq;
B1=rho1*etas[0]/sigma2sq+(a1*etas[2]-a1*a2*etas[0])/sigma3sq+(phis[1]*etas[3]-phis[0]*phis[1]*etas[2]-
      phis[1]*phis[2]*etas[0]+phis[2]*etas[4]-phis[0]*phis[2]*etas[3]-phis[1]*phis[2]*etas[2])/sigvsqq;
muet = B1/A1;
sigetsq = 1.0/A1;
siget =sqrt(sigetsq);
sigetprop=siget;
betxt=0.0;

// finding X^T*Beta

for(j=0;j<betnum;j++) betxt+=betas[j*gibit+i-1]*col[1][j];

ydat=col4[1];

// the mean of the envelop function : mean

nlm(&a,&betxt,&ydat,&muet,&sigetsq);

expbetxt=exp(betxt+a);

mean=muet+sigetsq*(ydat-expbetxt);

//## THE ACCEPT-REJECT METHOD
count=0.0;
do{
    etprop=rnorm(mean,siget);
    u=log(runif(0,1));
    ratio=ydat*etprop-exp(betxt+etprop)+dnorm(etprop,muet,siget,1)-((a-1)*expbetxt+
              (mean*mean-muet*muet)/(2*sigetsq)+dnorm(etprop,mean,siget,1));
    count+=1.0;
}while(u>ratio);

etas[1] = etprop;

     //################################
     //## SAMPLING ETA_t t=p+1,...,(n-p)
//mean and variance for \eta_t: t=p+1,...,n
// note that the loop starts from p not p+1, because in C we start from 0 not 1

A1=1.0+phis[0]*phis[0]+phis[1]*phis[1]+phis[2]*phis[2];
sigetsq=sigvsqq/A1;
siget=sqrt(sigetsq);
sigetprop=siget;

for(j=p;j<n;j++)
  {
     B1=phis[0]*etas[j-1]+phis[1]*etas[j-2]+phis[2]*etas[j-3]+phis[0]*etas[j+1]-phis[0]*phis[1]*etas[j-1]-
```

```
            phis[0]*phis[2]*etas[j-2]+phis[1]*etas[j+2]-phis[0]*phis[1]*etas[j+1]-phis[1]*phis[2]*etas[j-1]+
            phis[2]*etas[j+3]-phis[0]*phis[2]*etas[j+2]-phis[1]*phis[2]*etas[j+1];

            muet=B1/A1;
            betxt=0.0;

            // X^T*Beta
            for(k=0;k<betnum;k++)  betxt+=betas[k*gibit+i-1]*col[j][k];

            ydat=col4[j];

            // mean for the envelope

            nlm(&a,&betxt,&ydat,&muet,&sigetsq);

            expbetxt=exp(betxt+a);

            mean=muet+sigetsq*(ydat-expbetxt);
            count=0.0;

            //## THE ACCEPT-REJECT METHOD

            do{
                etprop=rnorm(mean,siget);
                u=log(runif(0,1));
                ratio=ydat*etprop-exp(betxt+etprop)+dnorm(etprop,muet,siget,1)-((a-1)*expbetxt+
                          (mean*mean-muet*muet)/(2*sigetsq)+dnorm(etprop,mean,siget,1));
                count+=1.0;
            }while(u>ratio);


           etas[j] = etprop;

          }


// sampling betas one by one

fb1=0;

// fb1 is the loglikelihood of Beta based on old values

for(j=0;j<n;j++)
{
    ssum1=0.0;
    for(l=0;l<betnum;l++)
        ssum1+=betas[l*gibit+i-1]*col[j][l];

    fb1+= col4[j]*ssum1-exp(ssum1+etas[j]);
}

for(k=0;k<betnum;k++)
{
    for(j=0;j<betnum;j++)
    {
        if(j>=k)
        {
            betasmh[j]=betas[j*gibit+i-1];
        }else{
            betasmh[j]=betas[j*gibit+i];
        }
    }
    // Sampling from the proposal
    mode=betasmh[k];

    nlmbet(&mode,&stdev,&betasmh,cols,col4,&etas,betnumber,ss,&k,&mubet,&sigbetsq);

    betprop=rnorm(mode,stdev);
```

```
            betasmh[k]=betprop;

            // fb2 is the loglikelihood of beta based on the new sampled value
            fb2=0;

            for(j=0;j<n;j++)
            {
                ssum2=0.0;
                for(l=0;l<betnum;l++)
                    ssum2+=betasmh[l]*col[j][l];

                fb2+= col4[j]*ssum2-exp(ssum2+etas[j]);
            }

            // the Metropolis-Hastings step

            ratio1=fb1-((betas[k*gibit+i-1]-mubet)*(betas[k*gibit+i-1]-mubet))/(2*sigbetsq)-
                       ((betprop-mode)*(betprop-mode))/(2*stdev*stdev);

            ratio2=fb2-((betprop-mubet)*(betprop-mubet))/(2*sigbetsq)-
                       ((betas[k*gibit+i-1]-mode)*(betas[k*gibit+i-1]-mode))/(2*stdev*stdev);

            ratio=ratio2-ratio1;
            unii=log(runif(0,1));

            if(unii<ratio)
            {
                betas[k*gibit+i]=betprop;
                fb1=fb2;
                acceptbet[k]+=1.0;
            }else{
                betas[k*gibit+i]=betas[k*gibit+i-1];
            }

        }

        // sampling rstar as a block

//        log likelihood based on old values
        rstloglik(&logdens,&etas, &gammas,&phis,ss, &p, &sigvsqq);
  // log prior based on old values
        rstlogpr(&logpr,&rstar,&p);

        ratio1=logdens[0]+logpr[0];
        gam1=logdens[1];

        // sampling from the proposal
        for(j=0;j<p;j++) rstarprop[j]=rnorm(rstar[j],sigprop);

        getphi(&phis, &rstarprop, &p);
        getgammas(&gammas,&phis,&sigvsqq);

 //       loglikelihood based on the new values

        rstloglik(&logdens,&etas, &gammas,&phis,ss, &p, &sigvsqq);

 //       log prior based on new values

        rstlogpr(&logpr,&rstarprop,&p);

        // the M-H step

        ratio2=logdens[0]+logpr[0];
        ratio=ratio2-ratio1;
        unii=log(runif(0,1));

        if(unii<ratio)
```

```
        {
            for(j=0;j<p;j++) arpar[j*gibit+i]=rstarprop[j];
            for(j=0;j<p;j++) rstar[j]=rstarprop[j];
            gam1=logdens[1]; // needed to sample \sigma^2_V to avoid recomputing it again
            acceptrst[0]+=1.0;
        }else{
            for(j=0;j<p;j++) arpar[j*gibit+i]=arpar[j*gibit+i-1];
        }

        // sampling sigmavsq

        sigvsqq=1/rgamma(alfa1,1/gam1);
        arpar[p*gibit+i]=sigvsqq;

       }
 //   printf("%f\n",times);

    PutRNGstate();
}

// newton raphson method to find the mode of the posterior of \eta_t for all t

void nlm(double *a,double *betxt, double *yt, double *muet, double *sigsqet)
{
    double expbetxt,fp1,fp2,fpp,ans=log(*yt+1.0);
    expbetxt=exp(*betxt);
    fp2=expbetxt*exp(ans)-*yt+(ans-*muet)/(*sigsqet);
    do{
        fp1=fp2;
        fpp=expbetxt*exp(ans)+1.0/(*sigsqet);

        ans=ans-fp1/fpp;
        fp2=expbetxt*exp(ans)-*yt+(ans-*muet)/(*sigsqet);
    }while(fabsf(fp2-fp1)>0.000001);

    *a=ans;
 //   *stdev=1/sqrt(expbetxt*exp(ans)+1.0/(*sigsqet));
}

// newton raphson method to find the mode of the posterior of Beta

void nlmbet(double *mode,double *stdev,double *bett,double *xdatt, double *ydatt, double *etas, int *betnum,
                    int *ss,int *kk, double *mubet, double *sigsqbet)
{
    int i,j,l,k,n=*ss;
    double betxt[n];
    double fp1,fp2,fpp,ans=log(*ydatt+1.0);
    double expbetxt,s1[3];

    sfunc(&s1,ss,betnum,xdatt,bett,ydatt,etas,kk);

    // printf("%f %f %f\n",s1[0],s1[1],s1[2]);

    fp2= s1[0]-s1[2]+(ans-(*mubet))/(*sigsqbet);
    do{
        fp1=fp2;
        fpp= s1[1]+1.0/(*sigsqbet);

        ans=ans-fp1/fpp;

        bett[*kk]=ans;
        sfunc(&s1,ss,betnum,xdatt,bett,ydatt,etas,kk);
        fp2= s1[0]-s1[2]+(ans-(*mubet))/(*sigsqbet);

    }while(fabsf(fp2-fp1)>0.000001);

    *mode=ans;
    *stdev=1.0/sqrt(s1[1]+1.0/(*sigsqbet));
```

```
}

// this function is to evaluate some summations needed for the derivative of log posterior of Beta

void sfunc(double *s1,int *ss,int *betnum, double *xdatt, double *bett, double *ydatt, double *etas,int *kk)
{
    int i,j,k,l,n=*ss;
    double expbetxt,betxt[n];

    for(j=0;j<n;j++)
    {
        betxt[j]=0.0;

        for(l=0;l<*betnum;l++)
            betxt[j]+=bett[l]*xdatt[l*n+j];

    }

    s1[0]=0.0,s1[1]=0.0,s1[2]=0.0;

    for(j=0;j<n;j++)
    {
        expbetxt=exp(betxt[j]+etas[j]);
        s1[0]+=xdatt[(*kk)*n+j]*expbetxt;
        s1[1]+=xdatt[(*kk)*n+j]*xdatt[(*kk)*n+j]*expbetxt;
        s1[2]+=ydatt[j]*xdatt[(*kk)*n+j];
    }
}

// to get phi from rstar

void getphi(double *phivec, double *rstvec, int *pp)
{
    int i,j,k,pval;
    pval=*pp;
    double exprst,yy[pval][pval],rvec[pval];
    for(i=0;i<pval;i++)
    {
        exprst=exp(rstvec[i]);
        rvec[i]=(exprst-1.0)/(exprst+1.0);
    }
    for(i=0;i<pval;i++) yy[i][i]=rvec[i];

    for(k=1;k<pval;k++)
    {
        for(i=0;i<=(k-1);i++)
        {
            yy[i][k]=yy[i][k-1]-rvec[k]*yy[k-i-1][k-1];
        }
    }
    for(i=0;i<pval;i++) phivec[i]=yy[i][pval-1];
}

// the components of the covariance matrix of the first p=2 latent variables
void getgammas(double *gammas,double *phis,double *sigvsq1)
 {
 double deno,sigvsqq=*sigvsq1;
 deno=1-phis[0]*phis[0]-phis[1]-phis[0]*phis[0]*phis[1]-phis[1]*phis[1]*(1-phis[1]);
 gammas[0]=(1-phis[1])*sigvsqq/deno;
 gammas[1]=phis[0]*sigvsqq/deno;
gammas[2]=(phis[0]*phis[0]+phis[1]*(1-phis[1]))*sigvsqq/deno;

 }

// the covariance matrix of the first p variables of the AR(p) process

void covmatrix(double *covmat, double *gammas, int *pval)
{
```

```
    int j,k,p=*pval;

    for(k=0;k<p;k++)
    {
        for(j=0;j<(p-k);j++) covmat[j*(p+1)+k]=gammas[k];
        if(k>0){for(j=(p-1);j>=k;j--) covmat[j*(p+1)-k]=gammas[k];}

    }
}

// this function is to find the Chelosky decomposition of the covariance matrix

void chol(double *a,int *dim)
{
    double rcond, z;
    int info;
    // F77_CALL(dpoco)(a,dim,dim,&rcond,&z,&info);
    F77_CALL(dpofa)(a,dim,dim,&info);
}

// to find the inverse and determinant of the covarince matrix

void invdet(double *a,int *dim,double *det)
{
    int j,k,p=*dim,job=11;
    chol(a,dim);
    F77_CALL(dpodi)(a,dim,dim,det,&job);

    det[0]=det[0]*pow(10,det[1]);
    for(k=0;k<=(p-2);k++)
    {
        for(j=(k+1);j<=(p-1);j++) a[k*p+j]=a[j*p+k];
    }
}

// the log likelihood of rstar

void rstloglik(double *logdens,double *etas, double *gammas,double *phis, int *ss, int *p, double *sigvsq)
{
    int i,j,pp=*p,n=*ss;
    double covmat[pp*pp],det[2];
    double mean,sumet=0.0;
    double etcov[pp],prodans=0;


    for(i=0;i<pp;i++) etcov[i]=0.0;

    for(i=pp;i<n;i++)
    {
        mean=0.0;
        for(j=0;j<pp;j++) mean+=phis[j]*etas[i-j-1];
        sumet+=(etas[i]-mean)*(etas[i]-mean);
    }

    covmatrix(&covmat,gammas,p);

 //   for(i=0;i<(pp*pp);i++) printf("%f ",covmat[i]);
 //   printf("\n");
    invdet(&covmat,p,&det);

 //   printf("rstloglik\n");

    for(i=0;i<pp;i++)
    {
        for(j=0;j<pp;j++) etcov[i]+=etas[j]*covmat[pp*i+j];
        etcov[i]=etcov[i]*etas[i];
    }
```

```
    for(j=0;j<pp;j++) prodans+=etcov[j];

    logdens[0]=-0.5*(log(det[0])+prodans+sumet/(*sigvsq));
    logdens[1]=0.5*(prodans*(*sigvsq)+sumet);
}

// log prior of rstar

void rstlogpr(double *logpr,double *rstar,int *p)
{
    int k;
    double kd,logexprst;
    logpr[0]=0.0;
    for(k=0;k<(*p);k++)
    {
        kd= (double) k;
        logexprst=log(exp(rstar[k])+1.0);
        logpr[0]+=(floor(0.5*(kd+2.0))-1.0)*(rstar[k]-logexprst)-floor(0.5*(kd+1.0))*logexprst+rstar[k]-2.0*logexprst;
    }
}
```